DE MONTFORT
UNIVERSITY

LEICESTER • BEDFORD
MILTON KEYNES

Digital Watermarking
and
Novel Security Devices

A thesis presented by

Miss Savita De Souza

to

Faculty of Computing Sciences and Engineering

in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in the subject of Covert Bar Coding

De Montfort University
Leicester
United Kingdom

September 2003

Abstract

This research is in the field of document security and the investigation of existing security devices. Security documents can be of two types; those that have commercial value, namely currency, articles of value and those that have legal value, namely, passports, identification cards etc. It is of vital importance in today's world, where fraud and counterfeiting are the order of the day, to protect any article or document of aesthetic importance and value. After the tragic events of September 11th 2001, security and the lack of security in every sense of the word is a much talked and researched subject. The literature review in Chapter Two will reveal that there are many security devices being currently used for various applications and each of them has many advantages and disadvantages. Devices such as magnetic cards, chip cards, biometric technology, optically variable devices, holograms and kinegrams and others have been described and their uses explained. Chapter Three is a continuation of literature review on Digital Watermarking and the available Digimarc technologies and applications in the field of security. Digimarc Corporation, based in Tualatin, Oregon, United States has been the pioneer in the field of Digital Watermarking. Microbar Security Limited, a subsidiary company of Durand Technology Limited, based at the Shrivenam 100 Business Park, Shrivenam, Oxfordshire, United Kingdom has developed its own watermarking technique, which has many advantages in comparison to the Digimarc Technology has also been described here.

In Chapter Four, an optical variable device, namely, the Lippmann Optical Variable Device, named after the nineteenth century scientist Gabriel Lippmann (1845-1921), has been studied in thorough detail. Much work has been done to prove it and to understand the principle behind the theory he proposed. Currently, this type of technique based on interferential photography can be applied as a unique security device on passports, identification cards etc. The principles and theory of Lippmann photography has been explained in considerable detail. Experimental work done using various methods and using different photographic materials has also been described along with the results obtained. The application that has been emphasized here, is for passports for which this research was conducted in collaboration with Holographic Dimensions, Florida.

Chapter Five, explains the principle of Covert Bar Coding which is the basis of the Microbar Watermarking Technique. Experimental results have been presented, these have been obtained using the MATLAB software.

Finally, Chapter Six gives some ideas that can be pursued for future work.

This work has produced two papers on Lippmann Photography and the mathematical evaluation of the emulsion co-authored by the author of the thesis. Abstracts from the papers have been used in this thesis. Also, it has produced two patents, one by 'Holographic Dimensions', Florida on the Lippmann Photography and second, 'Microbar Security Limited' on the 'Covert Bar Coding Technique'.

Acknowledgements

I wish to thank my supervisor Professor Jonathan Blackledge (Former Director of Mcleods Institute of Mathematical & Simulation Sciences, De Montfort University, Leicester) for his constant help and guidance throughout my studies.

Many thanks to Mr. Bruce Murray, Professor Gwynne Evans and Dr. Martin Turner for their useful suggestions.

I also wish to acknowledge Professor Hans Bjelkhagen and Professor Nicholas Phillips (Centre for Modern Optics, De Montfort University) for giving me the opportunity to work with them for the first two years of this research.

I thank my parents, Alvito and Sushila De Souza, to whom I dedicate this thesis, for giving me the chance to study in the United Kingdom and always being supportive of all my decisions.

Last but not the least, I wish to thank Ashish for always being there for me and helping me in every way, especially editing and proof reading this thesis.

Contents

List of Figures

List of Tables

Glossary

Terms Used:

AAMVA standard: The American Association of Motor Vehicle Administrators

ANSI standard: American National Standards Institute

ATM: Automated Teller Machines

CD: Compact Disc

CIE: Commission International d'Eclairage

CMC7: Type of MICR Barcode

DOVID: Diffractive Optically Variable Image Device

EBCOT: Embedded Block Coding with Optimized Truncation

ECMS: Electronic Copyright Management Systems

EPROM: Electronic Programmable Read Only Memory

FFT: Fast Fourier Transform

FRFT: Fractional Fourier Transformation

G_i: Gaussian distributed array

GP8: This is a developer used to develop the Lippmann photographs, for chemical composition, please refer to reference [93,94,102,103].

HOF: Higher Order Fractal

IC card: Integrated Circuit Card

IDC: Technology Market Analysts

IFS: Iterated Function System

InO: Indium Oxide

ISIS: Interference Security Image Structure

ISO standard: The International Organization for Standardization

LCD: Liquid Crystal Display

LSB: Least Significant Bit

MICR: Magnetic Ink Character Recognition

OCR: Optical Character Recognition

OID: Optically Invariable Devices

OVD: Optically Variable Device

OVG: Optically Variable Graphics

PC: Personal Computer

PDF47 standard: Two-Dimensional Type Barcode

PSDF: Power Spectrum Distribution Function

PSF: Point Spread Function

PVC: Polyvinyl Chloride

RSA: Rivest-Shamir-Aldeman Encryption

RSF: Random Scaling Fractal

SAR: Synthetic Aperture Radar

SFD: Stochastic Fractional Differential

USBC: United States Banknote Company

UV: Ultra-violet

WORM: Write Once, Read Many

Notations Used:

λ: wavelength of light in air

n: refractive index of the emulsion

Kelvin: Unit of value on the absolute temperature scale; 0^0 Kelvin (absolute zero) = -273.16^{0}C

PFG-03C: These are photographic plates used to record art holograms of red and green laser light, for demonstration in white light, made in Russia by SLAVICH Joint Stock Company

nm: nanometer used as a unit of emulsion thickness

lp/mm: lines per millimetre as a unit of resolution

J/cm^2: Joules per square centimetre used as a unit of sensitivity of RGB (Red, Green and Blue) Light

μm: micrometer used as a unit of emulsion thickness

HRF-700X071-3 film: This is a batch number of panchromatic photopolymer material manufactured by DuPont which has produced the first successful Lippmann photographs

Å: Degree Angstrom

KBr: Potassium Bromide

$AgNO_3$: Silver Nitrate

ml: millilitres

Ag_2S: Silver Sulphide

υ: frequency of light

δ: Unit used to denote length of Euclidean Curves

Chapter 1

Introduction

In the digital age, the need to rely on secure communications is an important matter not only for diplomatic or military purposes, but also for businesses such as e-commerce, marketing strategies etc. Even though Biometrics and other forms of authentication are generating a lot of interest, the high cost of these techniques is going to delay the mass adoption. Dataquest, a well known organisation which provides statistical information on Information Technology and telecom markets, cites a number of drivers for the increased security spending, from heightened awareness since the September 11th 2001 terrorist attacks to well-publicised hacks, virus outbreaks and distributed denial of service attacks that have crippled some web sites.

As rapidly as the enterprise world has become Internet enabled, so have the number of ways various security threats can attack networks, disable operations and compromise business integrity. The nature and complexity of threats is constantly changing, and a 'one threat, one cure approach' is clearly not enough. The bottom line is today's threats are becoming more devious and so the security with which we tackle them has to be smarter.

Securing our countries, our businesses, and our personal lives against

cyber crime and terrorism requires an unprecedented change in laws, policies, culture, and attitudes about security, which will evolve over time. As an example, we can see that the dawn of the new century revealed completely new threats to the United States. The tragic events of September 11th 2001, demonstrated that relatively low technology attacks can cause massive casualties, extensive social and business disruptions, and huge financial burdens. Therefore, we can see that security - or rather, lack of security has become a major issue for businesses and society.

The Figure 1.1 below shows that managed security revenues are expected to grow nearly 25% annually through 2005. The term security

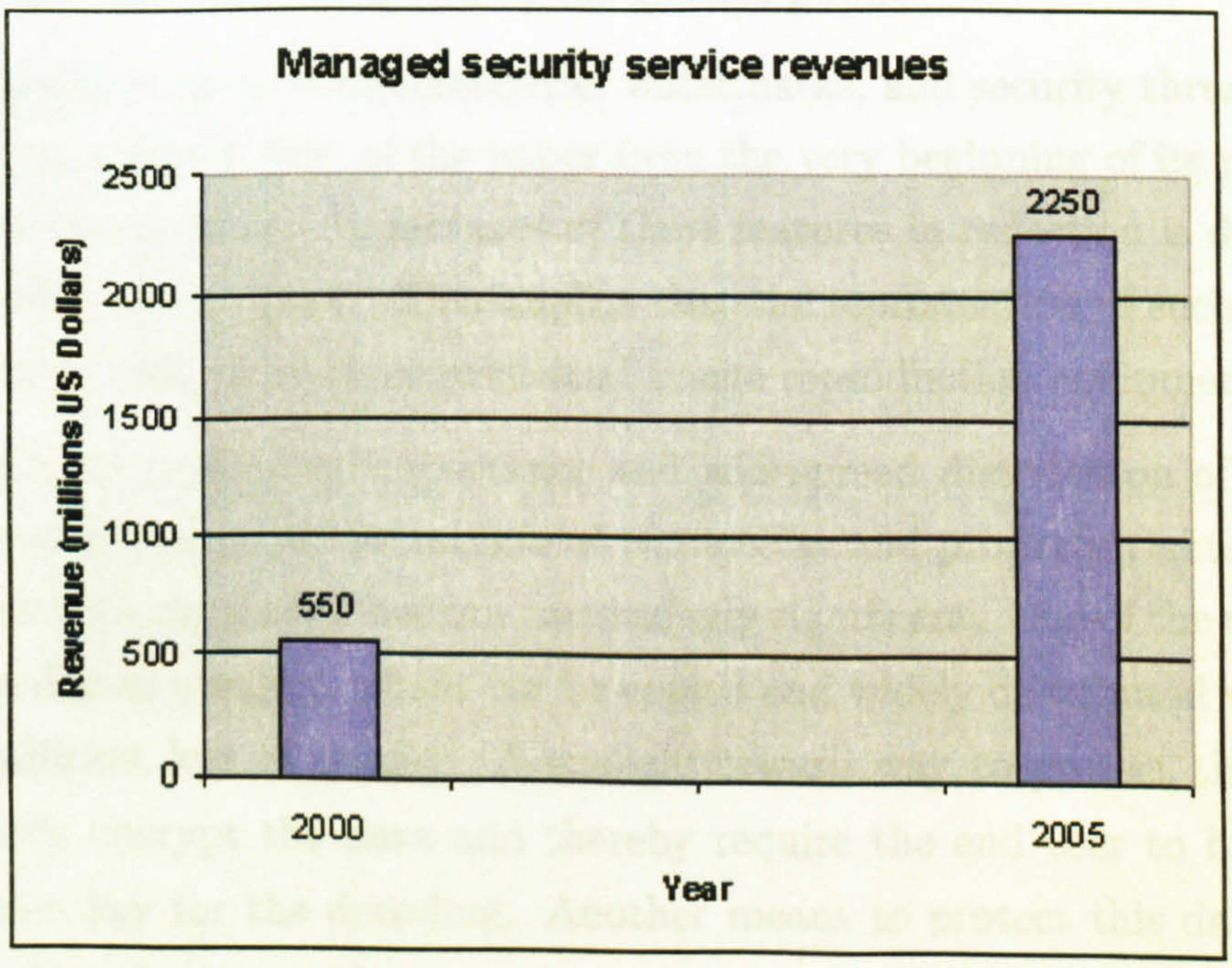


Figure 1.1: Managed security service revenues are expected to grow nearly 25% annually through 2005 [data for the graph has been obtained from Dataquest]

document refers to a printed medium, the authenticity and validity of which we can verify. Security documents may have commercial value (currency and cheques) or legal value (passports, ID cards). There are indeed two kinds of threats for a security document: counterfeiting and falsification.

A security document is realised by printing on paper an image that will help the public recognition of the document. Security printers use specific inks and patterns to differentiate the images produced from those used in commercial printings. However, due to the development of equipment such as colour copiers and scanners, which are commonly available and often require only a low level of skill to operate, it is easy to produce images of a good quality, allowing easy deception of the general public.

Features such as nonfluorescence, watermarks, and security threads can be made an integral part of the paper from the very beginning of its production. Furthermore, the appearance of these features in reflection is different from that in transmission. This implies that the reproduction of such effects is extremely difficult with conventional image reproduction equipment.

With the increasing importance and widespread distribution of digital media and imagery, the protection of the intellectual property rights of the owner for their media has become increasingly significant. One of the types of media is digital imagery, which can be copied and widely distributed without any significant loss of quality. A straightforward way to protect this is to completely encrypt the data and thereby require the end user to have the encryption key for the decoding. Another means to protect this data is to apply a digital watermark.

The 'watermark' has become synonymous with security because of its long and reliable use. It consists of an image integrated in a document that is visible chiefly in transmitted light and slightly distinguishable in reflected light. It is made during the wet paper-sheet formation and consists of varia-

tions in paper thickness and density, resulting in differences in sheet opacity. The watermark was probably the first OVD invented. These are various types of watermarks and many ways of embedding them in digital media.

A watermark may be produced by two different paper processes, applying either a Fourdrinier paper machine or a cylinder mould-made paper machine. The most secure watermark is the mould made watermark because it is more detailed and distinct, contains more visible halftones and has strong tactile relief.

To ensure a good security level, the watermark must be three dimensional, multitone with an accurate position within a precise area of the security document.

Digital watermarking is the process by which an image is coded with an owner's watermark and can be done using either of two general approaches. One approach is to transform the host image into its frequency domain representation and embed the watermark data therein. The second is to directly treat the spatial domain data of the host image to embed the watermark.

Regardless of the embedding method, there are several requirements that the embedding technique must satisfy. First, the watermarked image should retain as closely as possible the quality of the original image. This means that it should not be distinguishable that there is a watermark embedded in the image. Second, once the watermark has been recovered from the image, there should be a very high level of certainty as to whether the image was actually a watermarked image. Highly desirable also is that this determination can be made easily, yet accurately. Third, the watermark should be robust to various types of image processing techniques. This requirement is due to the common application of such techniques, most notably compression, as well as the possibility that these techniques may be applied with the intent to destroy the watermark in the image.

1.1 Background to the thesis

The current work describes the various security devices available and their many applications in the different fields and a literature review on digital watermarking techniques. Work has been done in the area of Lippmann photography, a one hundred year old technique, invented in 1891 by Gabriel Lippmann. Lippmann developed the first theory of recording monochromatic and polychromatic spectra. Currently, this type of technique can be applied as a unique security device on security documents, like passports, identification cards, credit cards, driving licences etc. This method offers a very high degree of security and has many advantages which have been investigated and experiments have been carried out to prove the same. Lippmann photographs are very similar to holograms currently used in security, but this technique gives a unique recording of each document so that a much higher degree of security can be achieved.

As mentioned earlier, digital watermarking can be used as a security feature wherein due to the embedded watermark, important documents or articles of value can be prevented from being counterfeited. The Microbar Covert Bar Coding Technique has been explained and experimentally proved. It was not possible to make a clear comparison between the technology used by the current leaders in this area, namely, Digimarc Corporation with respect to the watermarking algorithm used due to the technology being confidential to Digimarc. A brief overview of the Digimarc solution has been included in the literature review. The Microbar however, is based on the principles of fractal field and fractal geometry.

1.2 Structure of Thesis

The following chapter gives an overview of the main security devices currently on the market and their suitability for various applications. Devices such as laminates, smart cards, biometric technology and others have been considered along with their advantages and disadvantages. An up to date review on the various watermarking techniques currently in use has been presented in Chapter Three. Digimarc Corporation have until now, been the pioneers of this technique which involves embedding the Digimarc watermark in images for their protection against forgery. In Chapter Four, the Lippmann technology for passports and identification cards has been investigated. Many practical experiments have been conducted to achieve an almost perfect Lippmann photograph of various objects. A perfect reproduction is very difficult to achieve as it depends on many factors such as the emulsion used and its properties, time of development as well as the general conditions of the dark room. All the experiments were conducted in the Modern Optics Laboratory (Dark Room) at De Montfort University, Leicester. The results have been presented along with an analysis of the recording materials used. Chapter Five explains the Microbar technique based on the matched filter and linear frequency modulation principle and its suitability in watermarking. The software has been written using MATLAB. This technique can be further developed into a two dimensional watermark using the Fresnel theory of Optics and the principles of convolution and correlation. Chapter Six includes conclusions and further research directions in this field.

1.3 Original Contribution

The work described in the thesis is in the fields of optical security, namely the Lippmann photographic technique and digital security, namely the Microbar Covert Bar Coding and Texture Watermarking. The novelty in the Lippmann Technique was finding a new application for it in the field of security, as it cannot be copied by conventional photocopiers and scanners, hence its uniqueness. Even though this technique has been around since 1891, the first 18 months of the research were spent optimising the process of recording a Lippmann photograph and developing it to generate an optical variable device. The experiments have been conducted using the Lumiere and GP8 developers (see Chapter Four). With the facilities available at the Modern Optics Laboratory, De Montfort University, Leicester, photographic recordings were carried out under different conditions. In order to optimize the technique, various factors such as type of emulsion, size of grains in the emulsion, development time, recording time under various lighting conditions, etc. have been considered and varied accordingly to get the best results. Thus, the Lippmann photographic technique has been successfully studied and proved and its application to passports and identification cards has been practically implemented. Also the Lippmann photographs are very beautiful and have great archival stability. The Lippmann optical variable device (OVD) has very high resolution and is bragg sensitive, which means that the colour of the image recorded changes with the angle of illumination. This feature is very unique and has been successfully demonstrated. The author has also given suggestions to improve the method further.

At this point, the author moved on to another area in security, namely using fractals and chaos as a watermarking technique. The Matched Filter (see Chapter Five) has been used to develop the covert bar code which can be used for product individualisation. This application is vital as product

manufacturers lose as much revenue in counterfeit products as in product diversion. The author has been involved in experimenting with the prototype pen scanner, details of which can be found in Chapter Five. The idea of one-dimensional watermarking has been extended to the two-dimensional stage using Fresnel based techniques, a detail analysis of which has also been carried out. The author found that the number of Fresnel rings and signal-to-noise ratio are the two key factors in watermark recovery. The author has also contributed to the software development of the texture watermarking technique using fractals. The 'Microbar' is invisible to the naked eye and can be easily hidden in the background of an image or security documents. Microbar is far more secure than any other encryption method known, and from a mathematical point of view, it is effectively impossible to crack the coding system used.

Chapter 2

Literature Review A: Overview of Security Devices

2.1 The Evaluation of Document Fraud Resistance

This subsection focuses on security aspects of designs that aim at protecting valuable documents such as bank notes, ID cards, passports and products such as software, perfume, and watches to name a few examples. It is generally accepted that counterfeit resistance of security devices, and in particular diffractive OVD's, is an increasing function of their image complexity. Several hologram-manufacturing companies explicitly advocate the high image complexity of their products as an advantageous property that prevents counterfeiting. The following design aspects that must be taken into consideration when designing first-line security features and their functional integration in the overall design:

- The security feature must convey a message relevant to the product.

- The function of the feature must be obvious and intelligible.
- The functions must be standardised. If the functions are diverse and if the designs periodically change layout, they will not likely become understood and will be complex to use.

Figure 2.1 gives an overall picture of the design aspects involved in the design of a security feature.

Technical	**Economic**	**Environment**	**Planning**
Operands	Costs	***Human***	Timescale
Materials	Market Constraints	Health	Production Quantity
Size,weight	Company Constraints	Safety	Personnel
Shape,color	Competition	Recycling	Supplies
	Customer	Disposal	Storage
Properties		Facilities	Distribution
Security		Working Conditions	Training
Ergonomics			Documentation
Aesthetics			
Performance			
Quality			
Tolerances			
Reliability			
Durability			
Shelf Life			
Procedures			
Standardization			
Evaluation			
Maintenance			
Packing			
Installation			

Figure 2.1: Requirements for a security product

Fraud with valuable documents and products has two aspects.

Counterfeit: The reproduction of a document, article or security feature with the intent to deceive the close scrutiny of a qualified examiner.

Forgery: The replication or alteration of a document's data with the intent to defraud, for example a cheque amount, signature, data, etc.

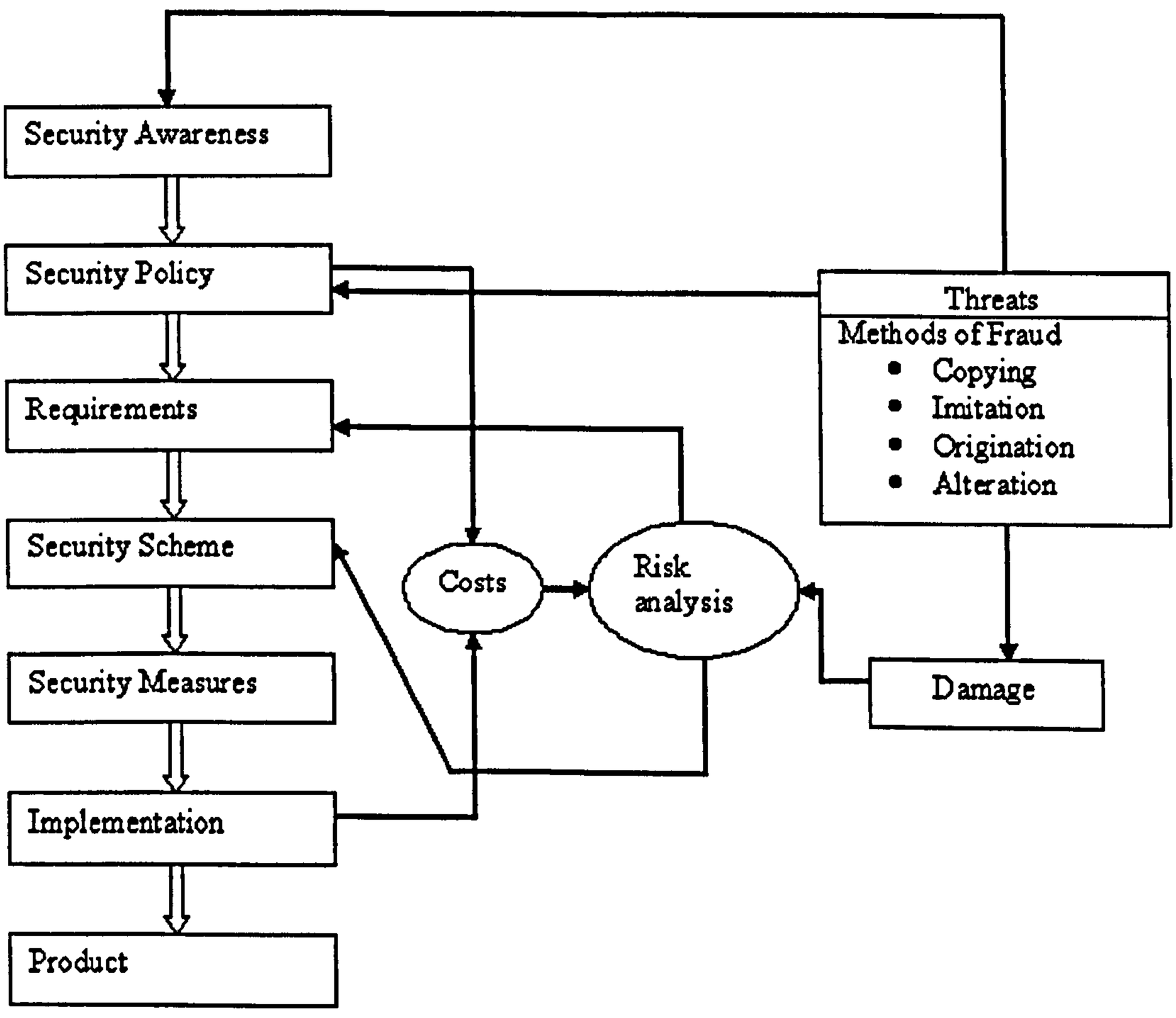


Figure 2.2: Development of a security product

In the case of documents that are not personalized, the aim of the counterfeiter could be to replicate all or parts of the document. Three approaches

are commonly used:

- reproduce a document using exactly the same materials and production methods used to create the original;
- replicate the document using like materials, production methods, or both; or
- replicate the document using entirely different materials and methods, to generate an imitation or look-alike of the original.

In addition to these methods, criminals might alter data on an unpersonalized document, perhaps to increase its original value.

Personalized documents can also be replicated in the same way as described above and subsequently altered for illegal use. A method of fraud that has been hard to protect until recently, against other than with rigorous inspection, is the use of genuine and unaltered passports by persons who sufficiently resemble the description data and the original holder depicted in the passport photograph like, family members.

Employing organizational security measures throughout the entire logistic path of the document is, an important way to limit origination[1]. This includes ensuring that equipment, materials, and products are protected during storage and transportation. It also involves putting in place measures that monitor the availability of equipment and materials to authorised parties only from the early stages of development.

The security measures that protect documents from being copied, imitated or altered are continually evolving as copying methods and other technologies evolve. An interesting aspect of this evolution is that new security measures in turn necessitate complementing security devices. For instance

the use of specialised laminates to discourage alteration of personalised documents requires a monitoring device to verify that the laminate has not been tampered with.

The choice of security features is determined by a benefit (weighing the benefits of fraud protection) versus cost (price of a complete security system) analysis. The following aspects should be considered when choosing an appropriate security system: What attack method is expected to cause the highest losses? Against which attack method is protection most needed and at what price? In any eventuality, attack systems that can shutdown/destroy the entire system must be avoided at any cost.

Ultimately, the effectiveness of the chosen security measures depends not only on the security features as such, but also on the overall security system, including the organisational measures accompanying the design, production, storage, distribution and the actual inspection method. In certain cases it could be beneficial to choose public security features, whereas in others, second-line features are needed that can only be inspected with specialised equipment.

2.2 Types of security devices

2.2.1 Tamper sensitive paper

Tamper sensitive papers are security papers that are treated to expose any attempt to alter the writing or printing on them[2]. It is the unique combination of its surface properties that permit high image quality and durable printing and its ability to incorporate security features that makes it the best support for security documents. For these reasons, it can offer a wide range of solutions to fight most known threats of counterfeiting and forgery.

2.2.2 Optical Security in Laminates

Materials or documents composed of several layers are manufactured by a laminating process and the products formed in this way are called laminates. The security features that can be obtained using laminates are as follows:

- The document can be protected against wear and the effects of various substances or vapours such as moisture.
- It can be reinforced so that creases or crinkles occur less rapidly.
- It should be obvious that data are less legible on crumpled documents and that document creases are apt to show signs of wear more quickly.
- Its appearance can be enhanced by providing it with a glossy surface.
- Making alterations or falsifications can be discouraged because inks are inaccessible to bleaches and mechanical damage is easily visible on a flat regular paper surface rather than on a more irregular surface.

- The extra layer makes the document more complex, adding special properties that make counterfeiting more complicated and difficult.

However, the drawbacks are as follows:

- The original surface structure of the document is lost. This is especially noticeable when the document has certain relief as the result of the printing technique applied, for example letterpress, intaglio, stamp printing. The top layer can be too thick to leave any visual or tangible sign of the original relief.

- The same applies to the effect on a watermark. This can only be visible as light and dark patches in the paper, no longer showing variations in thickness.

- If the quality of the protective layer is not of the correct standard, it might eventually cause discolouration in the document.

Formerly, covering layers consisted of lacquer or varnish. With the advent of synthetic and semi synthetic transparent foils, it became possible to apply protective layers on documents in a laminating process, which can be adapted to the users requirements. Compared to the application of the lacquer, the advantages of the laminating are that no drying time is required and the manufacturing process is much cleaner, that is there is no lacquer waste and soiled equipment and tools.

Types of laminates: The laminates used as security documents are of two types depending on their manufacturing method:

- Fused laminates: Two chemically similar materials are fused by heat and pressure, for example, a plastic credit card.

- Adhesive laminate: Two chemically different materials are joined by a separate adhesive. Examples include paper documents on which personal information is typed, laminated in a plastic pouch, like a passport sheet, containing photograph and personal information and covered by transparent plastic foil.

Materials such as linear polymers which are flexible and pliable, have relatively stable dimesions, strong and difficult to tear can be used[3].

Distinguishing characteristics of fused and adhesive laminating processes:

- The fused laminating process, which takes a relatively long time per card enables a large number of cards to be made simultaneously. It is applied by major card producers, linked with a centralised manufacturing process, who usually have the required equipment for providing these cards with the card holder's personalisation.

- The adhesive laminating process, which, albeit labour intensive for large quantities, can be performed quickly and easily for single cards. Hence, it is often applied in decentralised card manufacturing.

- The fused laminating process has undoubtedly better adhesive qualities than the adhesive process. Delamination of the composite layers is impossible; printing ink however, can be detrimental to the adhesive properties. In principle, it is possible to separate the adhesive laminates by melting the adhesive[3].

Plastic cards, to a large extent, carry nonvisual readable data. To enable the input of this data, the cards are provided with a magnetic stripe or an integrated circuit (IC card or a chip card). The magnetic stripe may be applied by means of foil print or it may be integrated in the laminating

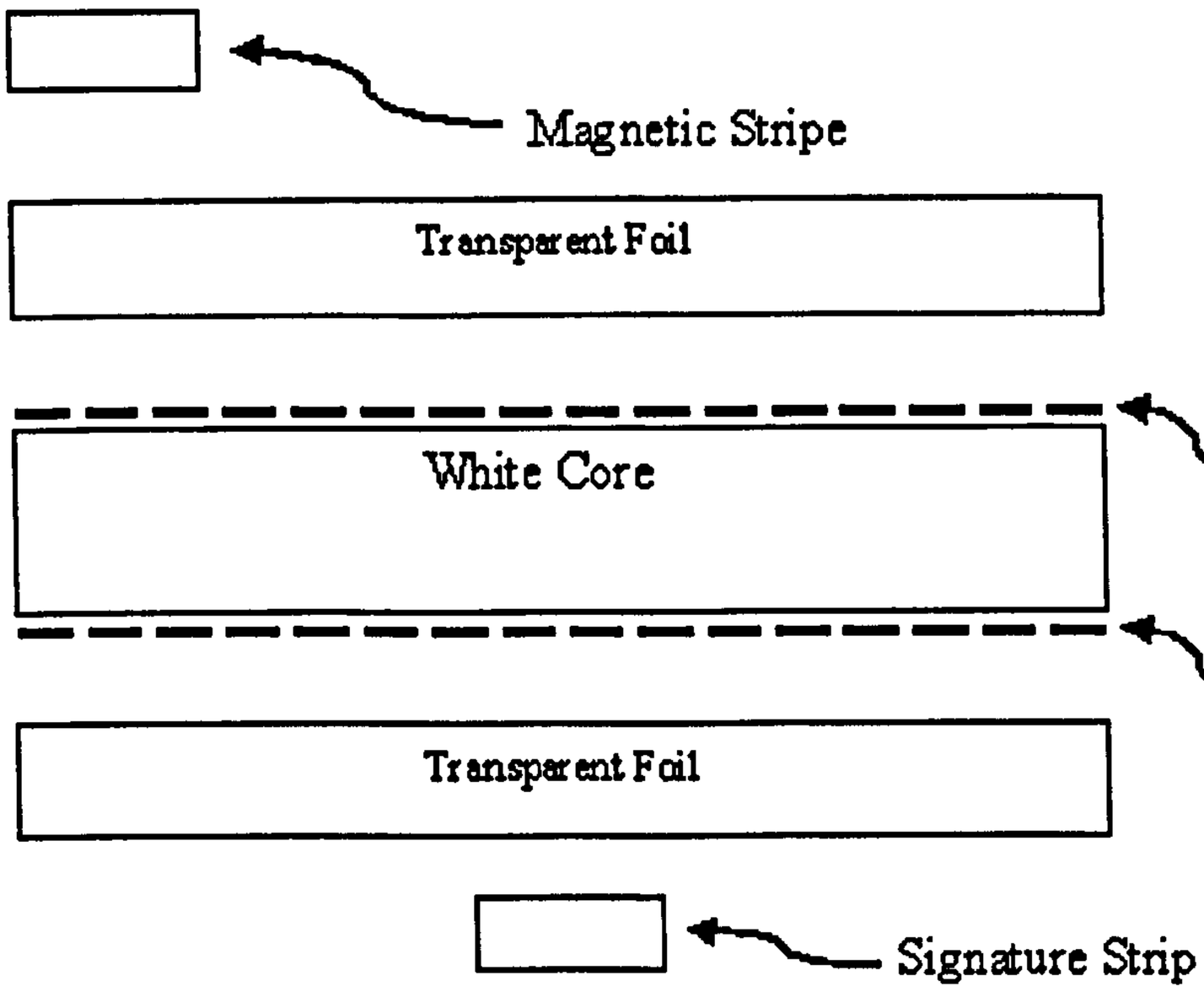


Figure 2.3: Magnified view of a possible composition of a laminated plastic card with signature strip and magnetic stripe

process. Another type of data carrying card is the optical memory card, which resembles the compact disc.

To discourage the reproduction of cards, they can be provided with OVD's, like embossed holograms. Semitransparent holograms are used to protect the underlying data and nontransparent reflection holograms can be used as carriers of visually legible data. Other diffractive OVD's such as exelgrams, kinegrams, or iridescent OVD's in general can also be used. In many applications, transparent holograms may be as large as the whole document itself in which case they are known as holographic foils[3].

Normally, cards are personalised by three main methods; the classical method being embossing the data. Another method is laser engraving of the card and a third method is thermal transfer printing, involving the use of

an ink ribbon containing a wax or resin-based ink. However, during recent years, new devices developed are connected to a control unit consisting of a computer system. The personal data of the prospective card holders is fed into the computer and stored. The control unit can be connected to a scanner or a video camera for scanning or taking a photograph and/or a signature or it can also be connected to a device for recording biometric characteristics.

2.2.3 Visual and Optical Character Recognition

Data can either be checked visually or by machine reading, referred to as optical character recognition (OCR). If the visual legibility is a main concern, then from a security viewpoint it will be advantageous to utilise a typeface that deviates from the types generally used. Machine readable characters can be read by both instrument and the naked eye, while, additionally, various optically readable codes or otherwise may be used. At present, OCR-B characters are most frequently used. Other types are OCR-A, magnetic ink character recognition (MICR), CMC7, and 7B for embossed cards[3]. Machine-readable data, which need not be visually legible, can be applied in the form of different types of codes like bar codes, dot codes, etc. These characters have a simple shape, essential for optimal readability, hence they do not provide any security, as reproduction can be fairly simple. Therefore, for enhanced security, such codes have to be covert. The technique of Covert Bar Coding has been fully investigated, implemented and has been described in Chapter Five.

Machine-readable documents are automatically inspected so that their data contents can be stored directly in data files to be processed and verified. But this type of data can often be easily copied or falsified, like in the case of normal bar codes. Hence, covert bar coding reduces the possibility or even eliminates the possibility of fraud and duplication.

Due to the fact that all these detection methods have become known widely even among counterfeiters, there is a growing demand for second line security features operating in accordance with these principles. By means of simple pocket equipment like hand-held scanners, UV sources and so on, a document can be quickly and unequivocally detected as authentic or fake. It can be said that all the ongoing research and investigations in the field of document security will always have the basic principles based on optics.

2.2.4 Non-Iridescent Optically Variable Devices

The expression optically variable devices (OVD) has been coined for iridescent security features that show various tilt effects, like kinematic effects and colour shifts, often obvious to the naked eye. Furthermore, OVD's cannot be copied by either colour copiers or colour scanners, and therefore demand great expertise from the would-be counterfeiter. It remains to be seen, however, to what extent these valuable properties are unique for iridescent security features. For this purpose, we must examine the main optical phenomena apart from diffraction and interference.

- *Diffuse Reflection* - paper, ink: white, black, colour, metameric colours;
- *Specular Reflection* - metallic foils, window thread;
- *Diffuse Transmission* - watermark, pseudowatermark;
- *Transmission* - windows in Australian plastic notes;
- *Tilt/Parallax* - transitory images, lenticular lens pattern devices;
- *Luminescence* - photosensitivity: fluorescence, phosphorescence;
- *Photochromism* - photosensitivity: colour changes activated by actinic light.

Except for the use of diffuse reflection of print, in general, these phenomena cannot be copied with the aid of current copiers.

2.2.5 Optically Variable Print

In general, diffusely reflecting security features, like microprint, fine-line patterns and intaglio print, are virtually invariant with respect to viewing angle and type of illumination (artificial light, daylight). They can therefore be considered optically invariable devices (OID's). Such OID's are most vulnerable to modern colour copying systems, and their major defence lies in creating fine detail near or below the resolution of such copiers[4].

It has been argued that the degree of micro structural order of optical security features is an approximate measure of their value for document security. Examples of highly ordered structures are interference security image structures (ISISs), such as multilayer thin films and liquid crystals, and diffractive optically variable image devices (DOVIDs) such as holograms, kinegrams, exelgrams and zero order devices. An entirely different type of security structure is the random structure, for example paper fibres and non-woven fibre structure, which combines a complete lack of structural order with a high degree of security. The mass production of random structures is relatively cheap compared to ordered structures. The reorigination of ordered structures is possible, while a random process never can be repeated and thus principally inhibits reorigination. Each random feature is unique per definition and are essentially limited to machine inspection.

Additionally, the non-woven structure is three-dimensional, hence impossible to counterfeit by any photographic reproduction technique. A few square millimetres of the non-woven structure are mounted in a visually or infrared transparent window in the security document, for instance a plastic

card or security label. The optical sensor captures two images of the structure under different viewing angles. Because the structure has a significant depth (approximately 0.3mm), the two images are distinctly different due to parallax of the fibres which serves as an authentication measure of the document[4]. Non-woven structures are extremely cheap and render high security, machine detectable devices.

2.2.6 Holograms as Anti-counterfeiting Labels

The first commercial use of embossed holograms in security was by United States Banknote Company (USBC) in 1980[5]. The embossed hologram is a good security device because it contains so much information that no single view of the image can include all the information contained by the hologram. It has to be viewed from several different angles to see the whole holographic object. A security document containing a hologram cannot be counterfeited by photography and printing without forsaking the parallax and changing colours.

The incorporation of a hologram into an otherwise complex, coherent document makes the hologram an extraordinary security feature. The success of the hologram as an authenticator of a security document rests not only on the difficulty surrounding its replication as a feature, but on the difficulty of incorporating the feature on or into the document itself.

A more recent development has been the integration of an optical element in smart cards. Generally thought to be very secure, nevertheless, they are exposed to a number of possible frauds including the immobilization of the chip and the replacement of the chip itself. One solution includes the machine verification of the optical signal included in the hologram, and the integration of that signal with the magnetic information. From these two

"numbers", either encrypted or simply arithmetic, the quotient of the two numbers is written to the EPROM of the smart card[5]. Whenever the card is swiped, the magnetic and holographic optical numbers are compared to the chip's number, and if they match the card is 'opened' for the transaction. The magnetic information on any card can be transferred, but the optical information effectively is hard wired directly into that particular card and cannot be changed. If the encrypted result of both magnetic and optical information matches the WORM part of the chip, it not only guarantees the authenticity of the chip but, should the chip be nonoperative for any reason, may allow (depending on the programming of the system) some partial transaction to take place (e.g. as a credit rather than a debit).

2.2.7 The Threat of Computers and Scanners

Recently, a way of desktop forgery has been developed that employs a workstation with high quality colour laser printer and scanner. As an example, say the object of the fraud is a legitimate bank cheque, made out to some company. The cheque is scanned into the computer. The signature and other information is left intact; only the numbers and the words are changed to a bigger amount. The manipulated image is printed out using the laser printer, using a magnetic ink toner for the black printing. Nowadays, this type of fraud is counteracted by manufacturing cheque paper that incorporates a holographic security stripe.

Unlike the credit cards, where only machine readable information is to be verified, there is a human component in the cheque processing chain. At the first human interface the person accepting the cheque does not know whether there are sufficient funds to cover the transaction. In many cases, banks have a signature book containing the signatures of all their customers and depositors and any cheque against their account for over a certain amount

is reviewed against the signature in the book. If the counterfeiter has not changed the signature on the falsified cheque, the bank will recognise the cheque as valid. However, it will be the holographic stripe that will trap the cheque writer.

2.2.8 Currency

Many countries such as Australia, Austria, Canada, Finland, Singapore, etc. have issued currency with OVD's. In the case of the Australian plastic bank note, the security feature itself is easily compromised by optical means, but the entire note is difficult to counterfeit because it involves the integration of the diffraction grating with a transparent plastic window and intaglio printing. There are three basic methods of using holograms in currency. The first approach is the hot stamping of a holographic feature on the finished note, for example using the portrait of the appropriate person, since human portraits are the most difficult type of holographic feature to simulate.

The second approach is perhaps the most cost effective: the holographic security thread. However, both these features are unable to survive extreme hostile environments and the hot stamping process is expensive. A better method is to cast the hologram, cure it with actinic radiation, and attach it to a polyester carrier. This creates a thread that can be serpentined into the paper, leaving the holographic windows at various points and creating a solid shadow in transmitted light[6].

The third approach, already in use is to incorporate a stripe of holographic foil at the leading edge of the currency note. The pattern is repetitive, thus relieving the necessity of registering an image to the paper. The distinctive holographic pattern can survive mutilation sufficiently to continue to identify the uniqueness of the note.

2.2.9 The Holomagnetic Stripe

The holomagnetic stripe consists of the combination of a holographic stripe with a diffractive code pattern and a magnetic stripe underneath called the Holomag^{TM} stripe. The light diffracted by the various diffractive patterns can be machine read. The optical signal thus obtained from the diffractive code pattern is subsequently linked with the traditional magnetic data, resulting in a card with extraordinary security. There is the option of storing the digitised optical information only at the host authorising computer, making it impossible for anyone in the field to know what the optical read was telling the host computer[5].

2.2.10 Optically Thin-film Security Devices

The effectiveness of optically thin-film security devices does not lie in the secrecy of their design, but in the lack of access of the public to materials with the same or similar iridescent properties. To prevent forgery using genuine security devices, it is important to ensure that they cannot be removed from a document and, after the document has been altered or replaced without the sign of tampering. This can be prevented by the incorporation of tamper-proof security seals or by the use of carefully selected adhesives and release agents which will not permit the intact removal of the coating from the document. To prevent counterfeiting, for example, by the transfer of an optically thin-film security device from a genuine lower denomination banknote to a higher denomination counterfeit, it is sufficient to customise the security device by the incorporation of a logo or denomination.

2.2.11 Additional Security Features

Patterns and logos introduced into the optically thin film security coatings significantly increase their effectiveness. They can be registered and legally protected by the methods available for the protection of trademarks and documents. The presence of logos prevents the use of stolen material on counterfeits of documents other than the one for which it was originally intended.

2.2.12 Liquid Crystal Displays (LCDs) & Document Security

An obvious application to optical document security is the incorporation of LCDs in documents. US patent number 4,472,627 describes a nematic LCD stacked with a thin solar cell as a photovoltaic device to drive the LCD[7]. Such a liquid crystal/photovoltaic device can be embodied in a plastic card with a transparent window or can be incorporated in paper documents like banknotes during the manufacture of the paper, whereby the paper fibres capture the device. Since the paper will be sufficiently translucent, the photovoltaic device will receive enough light to discharge the necessary voltage to drive the LCD.

2.2.13 Biometrics

Biometric technology is currently finding use in government applications, e-business, healthcare and computer networks all over. Instead of logging onto their networks with user names and the typical easily cracked passwords, about 2000 city employees in Glendale, California now enlist their fingerprints to securely sign on to their PCs[8]. In Belgium, the parliament used an

LCI-SMARTpen to sign an agreement with one of the government body's suppliers to authenticate prospective signatures. Though the SMARTpen device looks like an ordinary ballpoint pen, it contains sensors that capture the characteristics of a person's signature, like writing speed and pressure to authenticate proper identity.

Biometrics is by far the best technology to overcome the weaknesses associated with user name and password because it verifies human traits that cannot be passed to unauthorised users. The emerging biometrics industry will soon add a higher level of security and convenience to e-commerce transactions and user logon authentication applications. Combining biometric technology with public key infrastructure will also serve to bolster its potential applications.

2.2.14 Smart Cards

Smart cards have the advantages of simple deployment, ease of use, flexibility and smart because of their embedded integrated circuits. Information on a smart card can be divided into four main sections: information that is read only, added only, updated only and information with no access available. A smart card can restrict the use of information to an authorised person with a password. Some smart cards are capable of ciphering and deciphering so that the stored information can be transmitted without compromising confidentiality. They can cipher into billions and billions of foreign languages, and choose a different language at random every time they communicate. This authentication process ensures that only genuine cards and computers are used and makes eavesdropping virtually impossible, making this form of encryption powerful, effective, inexpensive, easy to use and providing information with confidentiality and integrity.

A cryptographic smart card features a microprocessor with special circuitry to quickly perform complex mathematical calculations, enough memory to store multiple digital credentials and perform all digital signature and encryption functions on the smart card itself[8]. By generating and storing private keys directly on the card, the keys are never present in the vulnerable desktop and server computer systems where they could be accessed or stolen. Only smart cards offer the ability to store a user's digital identity on a device that is both portable and powerful.

2.2.15 Printing Security

Designers of good security documents utilize three printing processes: Lithography, Intaglio and Letterpress. Each of these processes provides a unique benefit against the various forms of counterfeiting and combinations of these provide increased security.

- *Lithography*: Security printers use lithography in a fundamentally different way from commercial litho printers. In commercial printing, a total pictorial reproduction is achieved by breaking the picture into dots using the four-colour halftone process[9]. Security printers create images by using a line structure formation and changing individual colours along the length of the line. Simulated 3-D effects are achieved through combinations of overlaying special line patterns on separate print workings.

- *Intaglio*: Intaglio refers to print produced from plates on which the image is recessed. The recess is filled with ink, the surface cleaned and the ink transferred directly to the substrate. For security printing, thick paste inks are used with hand-engraved plates using high printing pressures. This gives security printing the distinctive variation in depth

of image and thus variation in tonal range. The pressure also creates associated embossing of the substrate. These variations in depth and image cannot be reproduced by planar printing processes such as litho, letterpress, etc[9].

- *Letterpress*: Security printers today still utilise letterpress printing to give each document an individual identification, principally achieved through the use of numbering. The counterfeiter has to choose between providing all documents with a similar number and hence aiding detection, or finding a way to reproduce a variable number output for each individual counterfeit.

2.2.16 Optically Variable Graphics Elements (OVG)

OVG elements are available with full-surface or partial metallization, full-surface transparent dielectric coating, and combinations of both. In applications such as driver's licenses and police ID cards, which often have to be verified under unfavourable lighting conditions, the kinegram is ideal. This is attributable to the specific kinegram microstructure that efficiently diffracts light in the direction of the observer.

Semitransparent OVG elements are available as see-through devices to cover information like ID photographs as a security protection. Because of their semi-transparency, see-through diffractive devices in general reflect considerably less light to the observer than fully metallised diffractive devices[10].

2.2.17 Zero-Order Grating Microstructures

Zero-order grating microstructures constitute a new class of security features with improved security against optical and physical copying. They exhibit unique optical behaviour suitable for both visual and machine inspection; since they are viewed in the zero order, they are visible even under adverse, diffuse illumination conditions. The microstructures can be fabricated by existing embossing and evaporation technology and can be fully laminated within plastic cards or in thin plastic film for application to paper[11]. They are designed with optical behaviour that is determined by the physical distribution of the refractive index within the grating microstructure, typically utilising a dielectric grating material with a high refractive index. This gives a very high degree of inherent security against forgery by optical copying techniques, for example, contact or holographic copying, since the optical effect can only be reproduced by an identical physical structure and not simply by recording the wavefront in a different medium.

2.2.18 Optically Variable Devices (OVDs)

Advantages of these devices has been discussed in Chapter Four on the Lippmann Optical Variable Device. The Figure 2.4 shows the various types of optical variable devices.

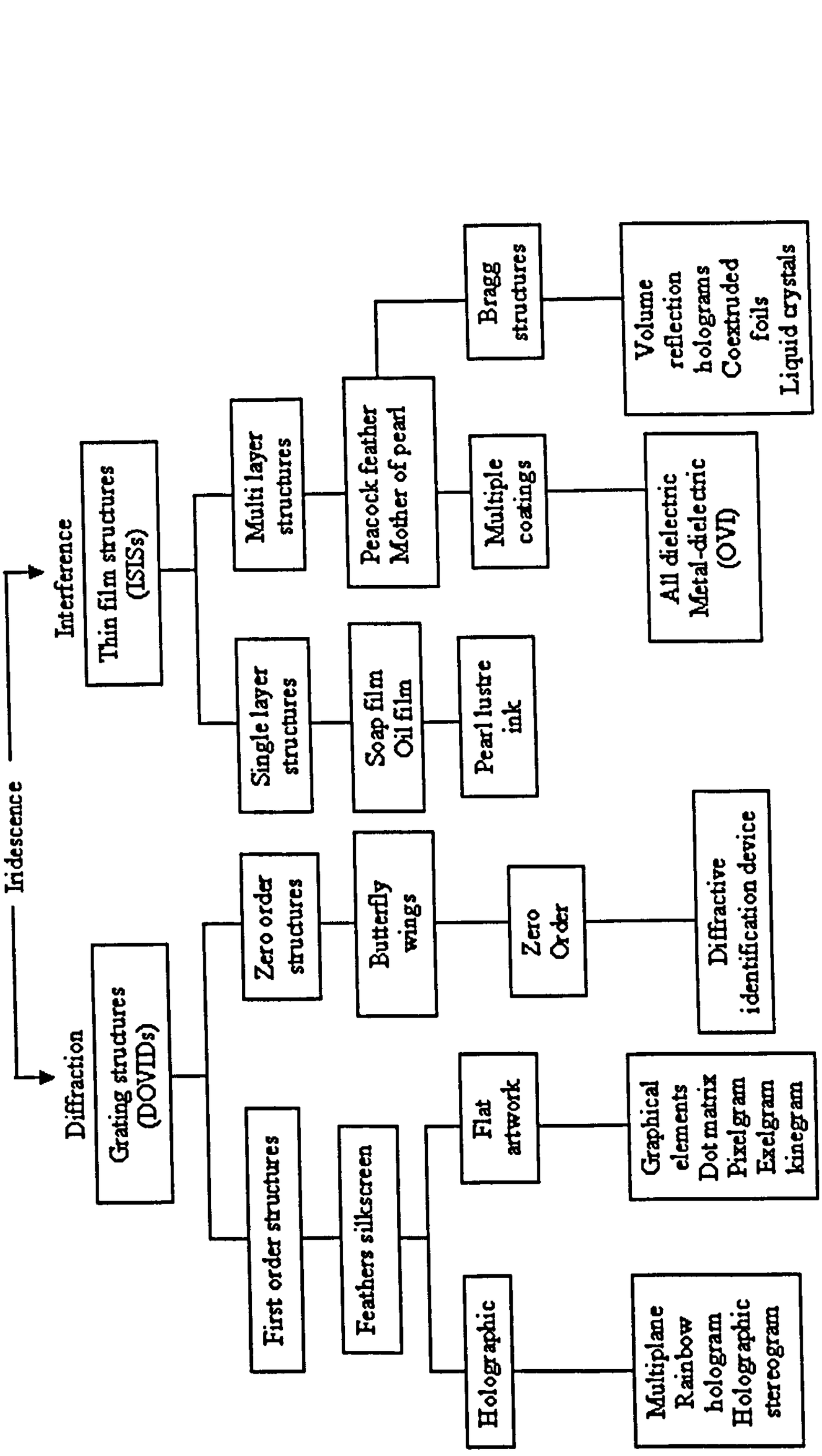


Figure 2.4: The tree of iridescent optically variable devices : iridescence in nature and as applied to document security

2.2.19 Retroreflective Security Devices

Most typical uses of the retroreflective products are for highway and vehicle reflectors, signs, license plates, warning emblems, and sports clothing. However, for the overlay films on the surfaces of official documents, much less retroreflectivity is required. Transparency is an essential feature to inspect when the document is covered with the retroreflective security film[12]. Any attempt to change a document protected by this overlay film involves the removal of or cutting through the film which breaks the continuity of the retroreflective image that is easily detected by inspection with either a hand held or table top verifier. Printing or marking on the surface of a retroreflector also destroys the retroreflective image. This product finds applications in passports, identification cards, visas, driver's licences, immigration cards, etc.

2.2.20 Types of Identification Cards

Magnetic Stripe Cards: A magnetic stripe is a half inch strip of magnetic tape, bonded to the card substrate, and encoded by aligning the magnetic particles along the tape in one direction or the other. Magnetic stripes can be read in swipe readers, insertion readers and motorised readers where either the card is moved past the head or vice versa. Magnetic stripe technology is reliable and very cheap. Its disadvantage is that it is now so well understood that cards can be very easily copied.

Chip cards: Chip cards can be either simple memory cards or complex microprocessor cards. These are much more difficult to counterfeit than the magnetic stripe cards thus lowering the opportunities for fraud. A chip card contains a single integrated circuit (IC) within the thickness of the card, connected to the outside world either by a set of contacts on the surface of the

card or by an antenna wound inside it. The IC's used range from very simple memory devices to 32 bit microprocessors with very special cryptographic coprocessors. The data stored in a chip card can be protected by a PIN number or a password.

Optical cards: These use broadly the same technology as the compact disc (CD), in which lasers are used to burn holes in a reflecting surface and detect the presence of these holes. These cards have a large non-volatile storage capacity as compared to other cards which makes them suitable for storing health records and other applications where a portable data carrier is required[8]. Although the data may be stored on the card in encrypted form, this largely defeats the purpose of making the data itself portable, since it means that the reader must always be connected on-line to a central system where the reference keys are stored.

Others: Other forms of machine-readable identification used on cards include the following:

- Bar codes: For simple and read-only identification, bar codes are one of the most reliable and cost-effective tools. The limitation of the technology is that it is read-only, although bar codes can be printed easily in the field and most codes can be copied and modified using computer printers and photocopiers. Copying of bar codes that are read in infra red can be thwarted by printing the code with carbon-based inks on a background that is dark or even black in the visible spectrum but transparent in infra-red[8]. An alternative is to print the code in a coloured ink that is sufficiently infra-red absorptive on a background of the same colour and that does not absorb infra-red.

- Two-dimensional bar codes: They offer much higher information density, and are more difficult to reproduce since they cannot be printed

by standard dot matrix or thermal printers, or by bar code software. These must be produced to high standards of linearity and rectangularity, and the absence of widely available hardware and software makes them more expensive than the common bar codes.

- Watermark tape: This can be used to enhance the security of the magnetic stripe cards. The tape is encoded with a unique code by its manufacturer and this code, or data derived from it, is also carried on the magnetic stripe. The proprietary process for producing the tape is kept secret and is assumed not to be reproducible.

- Holomagnetics: Here unique machine-readable data is contained in the hologram mounted on the card. This data is read by a proprietary device in an ATM or similar device and compared with the data on the magnetic stripe.

2.3 Conclusion

2.3.1 Security Effectiveness

The technical security features applied to documents must be consistent and effective against attack. It is not effective security if high-quality security paper permits written entries to be mechanically erased without leaving clearly distinguishable marks showing illicit alteration. Similarly, security effectiveness is reduced if the photograph on a high quality document can be substituted because of lack of attention to secure ways to affix the picture. Nor is it effective security if laminates can be removed from a document without destroying the underlying paper.

According to forensic statistics:

- Thin films and laminates should be more than just an adhesive film to cover a credit card or an ID document. Laminates with UV-luminescent printing, embossing structures and incorporating optically variable images or laser engraved information prevent simple replacement of the transparent film.

- Visa stickers with incorporated security features are more effective than traditional rubber stamps.

- Documents that incorporate photographs as an integral part of their material perhaps solve the problem of photograph substitution.

- Signatures of the holder of the document may be an useful tool for authentication. However, the signature must be authentic; scanned and reproduced signatures are of no value.

The key to success in detecting fraudulent security documents is the general inspection; consequently inspection personnel must receive comprehensive training. There are more and there are better counterfeited security documents demanding a concerted effort to document issuing authorities, document manufacturers, and law enforcement agencies to work together to increase the barriers for counterfeiting. Because security document counterfeiting involves organised crime, sometimes international strategies are required. Counterfeited and forged security documents have reached a quality where often the differentiation between counterfeit and genuine has become a real problem.

Chapter 3

Literature Review B: Digital Watermarking

3.1 Introduction

Digital Watermarking has recently become a popular research area due to the proliferation of digital data in this Internet age and the need to find a solution to protect the copyright of these materials.

Digimarc Corporation based in Tualatin, Oregon, United Sates has been the leading provider of digital watermarking components and technologies used in a wide range of security, identification and brand protection applications. Digital watermarking places an inherent digital identity into all media content, allowing media owners or issuers to: verify content, authenticate content, link media to related databases, manage and track digital assets, and prevent unauthorized copying.

This chapter is a literature review on available digital watermarking techniques, followed by a brief account of Digimarc and Microbar solutions.

What is watermarking?

Watermarking describes techniques which are used to convey information in a hidden manner by embedding the information into some 'innocent-looking' cover data. Watermarks are usually imperceptible, but there are also visible watermarks. Typically, this information is required to be robust against intentional removal by malicious parties. In contrast to cryptography, where the existence but not the meaning of the information is known, watermarking aims to hide the existence of the information from any potential eavesdropper altogether. Watermarking has existed since approximately the 15th century and in the past watermarks were mainly used on papers to identify the mill who made them. These are called physical watermarks as they exist in physical media. Nowadays, physical watermarks are commonly used to authenticate important documents, for example, banknotes and passports. With the advance of the Internet and the ubiquity of digital data, it is natural to extend the idea of watermarking to digital data. A popular application of digital watermark is to give proof of ownership of a piece of digital data (image, audio or video) by embedding copyright information as a watermark into the data itself.

Watermarking is closely related to steganography in that they are both concerned with covert communication and belong to a broader subject known as information hiding. Watermarking, is usually a one-to-many communication and has the added notion that the hidden message should be robust to attempts aimed at removing it. In the case of copyright protection, obviously the copyright information should resist any modifications by pirates intending to remove it.

There exists a duality between watermarking and data compression. While compression aims to identify the perceptually insignificant parts of the data and remove them, information hiding techniques try to insert informa-

tion into them. From an information-theoretical point of view, information hiding is a game between the information hider and the attacker. Moreover, compression is one of the most common operations on images, therefore one must take into account of the effects of compression when designing a watermarking system. The most common compression standard at the moment is JPEG, which is based on the discrete cosine transform (DCT).

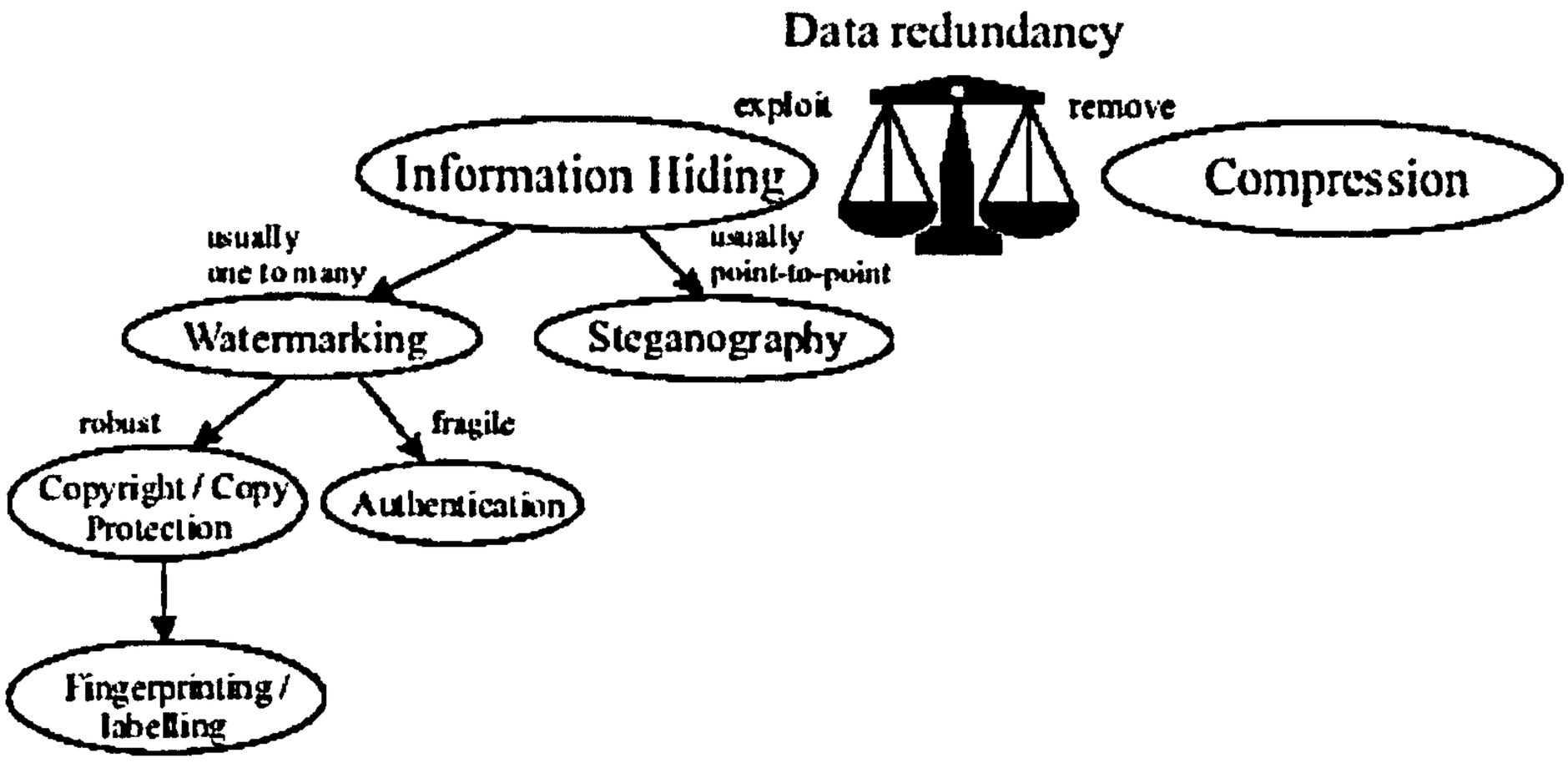


Figure 3.1: Relationship between digital watermarking, steganography, information hiding and compression[13].

3.2 Basic Watermarking Systems

All watermarking systems consist of an embedding part and a recovery part, which are shown in the following Figure 3.2.

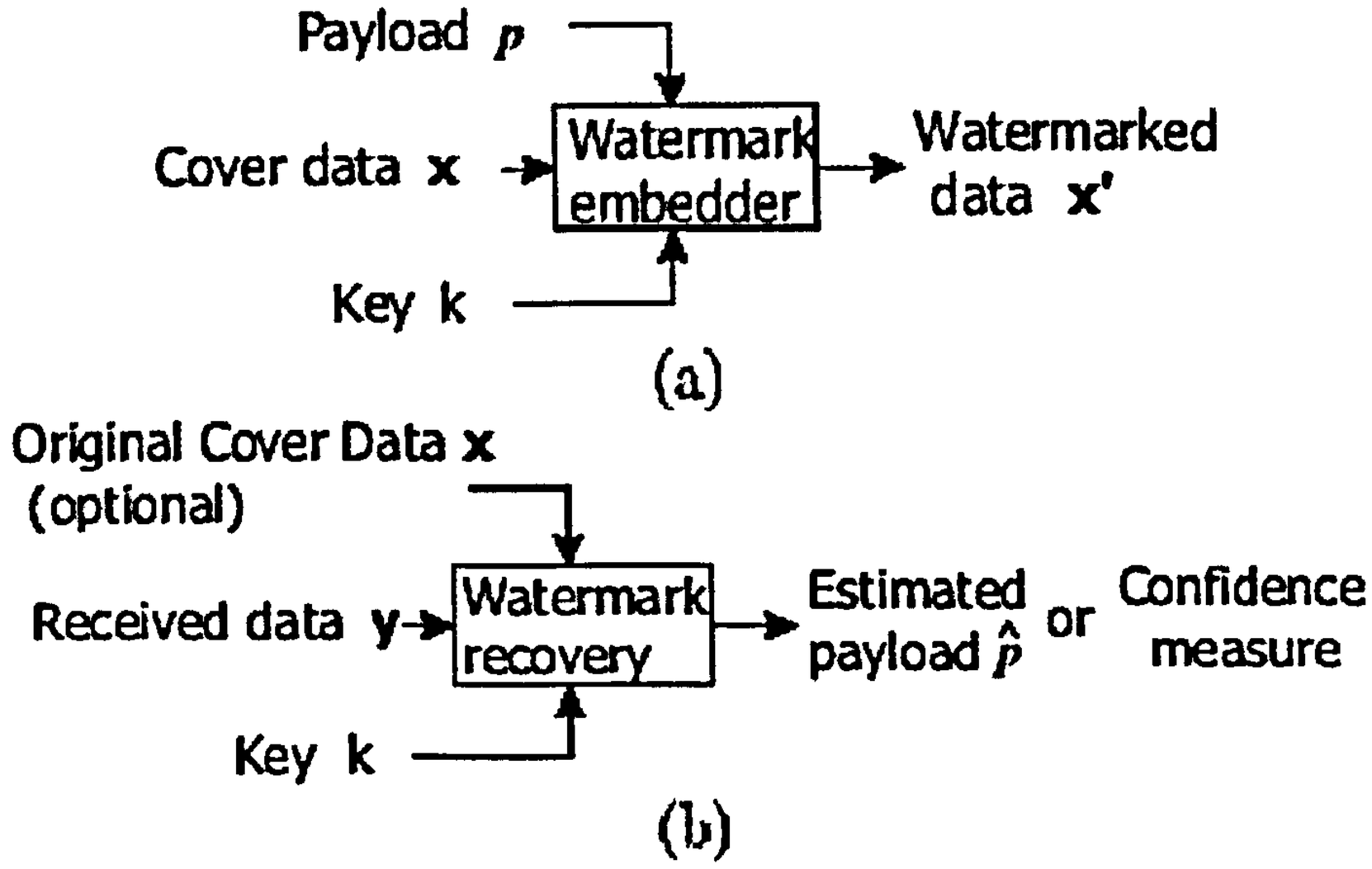


Figure 3.2: A generic watermark embedding (a) and recovery (b) system.

The embedder takes the cover data, the payload and a (public/secret) key to produce the watermarked data. The recovery part takes the (possibly modified) watermarked data, the key and/or original unwatermarked data and returns either the payload (decoding) or a confidence measure of how likely a specific watermark is present (detection)[14]. Regardless of the medium we are watermarking in, most watermarking systems share the following properties:

- **Imperceptibility:** The modifications caused by the watermarking system should be unobtrusive (with the exception of visible watermark obviously). This means one should use some sort of perceptual model, both for embedding a watermark and for evaluating the induced distortion. Being imperceptible also implies watermarks typically have much less power than the cover data.

- **Redundancy:** Since a watermark has much less power than its host, the watermark is usually redundantly embedded in the host to achieve robustness so that the payload can be recovered from a fraction of the watermarked data.

- **The use of keys:** All watermarking systems use one or more keys to ensure security against intentional removal of the hidden information. This is in accordance with the Kerckhoffs Principle in cryptography, in which the attacker is assumed to have full knowledge of the system and so security must lie only in the choice of key. If the watermark can be read, it may also be removed by the same person. This is why public watermarking still remains as a major challenge to researchers.

3.2.1 Watermarking Applications

Depending on the specific application of a watermarking system, the actual requirements will vary. Below are some of the applications of Watermarking.

- **Video Watermarking:** Watermarks can be created either in the spatial or in the DCT domains. In the latter, the results can be directly extrapolated to MPEG-2 sequences, although different actions must be taken for the I, P and B frames.

- **Audio Watermarking:** Here the time and frequency masking properties of the human ear are used to conceal the watermark and make it inaudible. The major problem lies in the difficulty in synchronising the watermark and the watermarked audio file but there are methods to overcome this problem which are beyond the scope of this thesis.

- **Text Watermarking:** At the printout level, information can be encoded in the way the textlines or words are separated which facilitates the survival of the watermark even to photocopying. At the semantic level, equivalences between words or expressions can be used.

- **Labeling:** The hidden message could also contain labels that allow annotating images or audio. With watermarking, it makes it difficult to destroy the label. An application is in medicine helping to prevent dangerous errors.

- **Fingerprinting:** The allows acquisition devices like video cameras, audio recorders etc. to insert information about the specific device like an ID number and date of creation. Watermarking makes it quite difficult to alter the signature though this feature can be achieved by other techniques.

- **Copy and Playback Control:** If the message carried by the watermark contains information regarding copy and display permissions then a secure module can be added to the copy or playback equipment to automatically extract this information and block further processing if required. This therefore requires agreements between content providers and consumer electronics manufacturers to introduce compliant watermark detectors in the video players and recorders. This is being used currently in Digital Video Discs (DVDs)[16].

- **Signalling:** The imperceptibility constraint is useful when transmitting signalling information in the hidden channel with the added advantage that no bandwidth increase is required. An application in broadcasting consists in watermarking commercials with signalling information that permits in automatic counting device to assess the number of times that the commercial has been broadcast during a certain

period.

- **Copyright Protection:** This is one of the major forces which drive the research in watermarking. Data can now be distributed in digital format with ease due to the existence of the Internet. The objective here is to embed copyright information into the data so that the rightful owner of a piece of data can at least prove his/her ownership in the case of a dispute[15]. The watermarks in this case obviously require a high level of robustness and should resist attempts in removing them. Watermarks for copyright protection do not prevent people from copying the digital data, they simply exist as a means for owners to assert ownership over some digital data.

- **Copy Protection:** In contrast to copyright protection, a copy protection mechanism actually prevents users from making unauthorised copies of the digital data. This is difficult in open systems like the Internet but it is possible to enforce copy protection in a controlled system like the DVD player. For example, the watermark which exists on a DVD, tells a compliant DVD player whether a user is allowed to copy the video.

- **Fingerprinting for pirate tracing:** Watermarks are used in fingerprinting applications to identify the legal recipient of the digital data and are typically used together with copyright protection watermarks in a transaction. The existence of multiple differently watermarked copies of the same data allow the collusion attack and so fingerprints must be designed to be collusion-secure. Watermarks for fingerprinting otherwise have identical requirements to that of copyright watermarks.

- **Watermarking for authentication:** Fragile watermarks are used to authenticate digital data. For example, if a digital photograph is to be

used as evidence in court, we have to prove that the photo has not been manipulated. A fragile watermark can be inserted into the image as soon as it is taken. If the image is modified maliciously, the watermark will be destroyed. If the watermark can be retrieved by the recipient, the image is deemed authentic, otherwise, it should be discarded as fake. A low level of compression is usually permitted but not content alteration of the image. Therefore a fragile watermark will have some robustness rather than like a checksum which fails even if only 1 bit of the data has been changed. In addition, it should be difficult for a malicious user to simultaneously modify both the cover data and the fragile watermark. There are several types of fragile watermarks; some only allow to detect if an image has been modified; some allow to calculate an approximate version of the original image in the modified regions; while others allow to invert the watermarking process and recover the original unwatermarked image if it is successfully authenticated. The last type finds applications in the medical industry as medical images cannot afford to be modified as it can cause misdiagnosis. Invertible fragile watermarks allow to both authenticate and recover the original digital medical image.

3.3 Current Watermarking Techniques

Since the publication of a seminal work by Tanaka et al. in 1990[18], there has been an extraordinary growth of techniques for copyright protection of different types of data, especially multimedia information. This interest is not surprising in view of the simplicity of digital copying and dissemination: digital copies can be made identical to the original and later reused or even manipulated.

There is an increasing need for software or in the worst case hardware, that allows for protection of ownership rights, and it is in this context that watermarking techniques become useful. Perceptible marks of ownership or authenticity have been around for centuries in the form of stamps, seals, signatures or classical watermarks, nevertheless, given current data manipulation technologies, imperceptible digital watermarks are mandatory for most applications. A digital watermark is a distinguishing piece of information that is adhered to the data that it is intended to protect, meaning that it should be very difficult to extract or remove the watermark from the watermarked image. Since watermarking can be applied to various types of data, the imperceptibility constraint will take different forms, depending on the properties of the recipient i.e. the human senses in most practical cases.

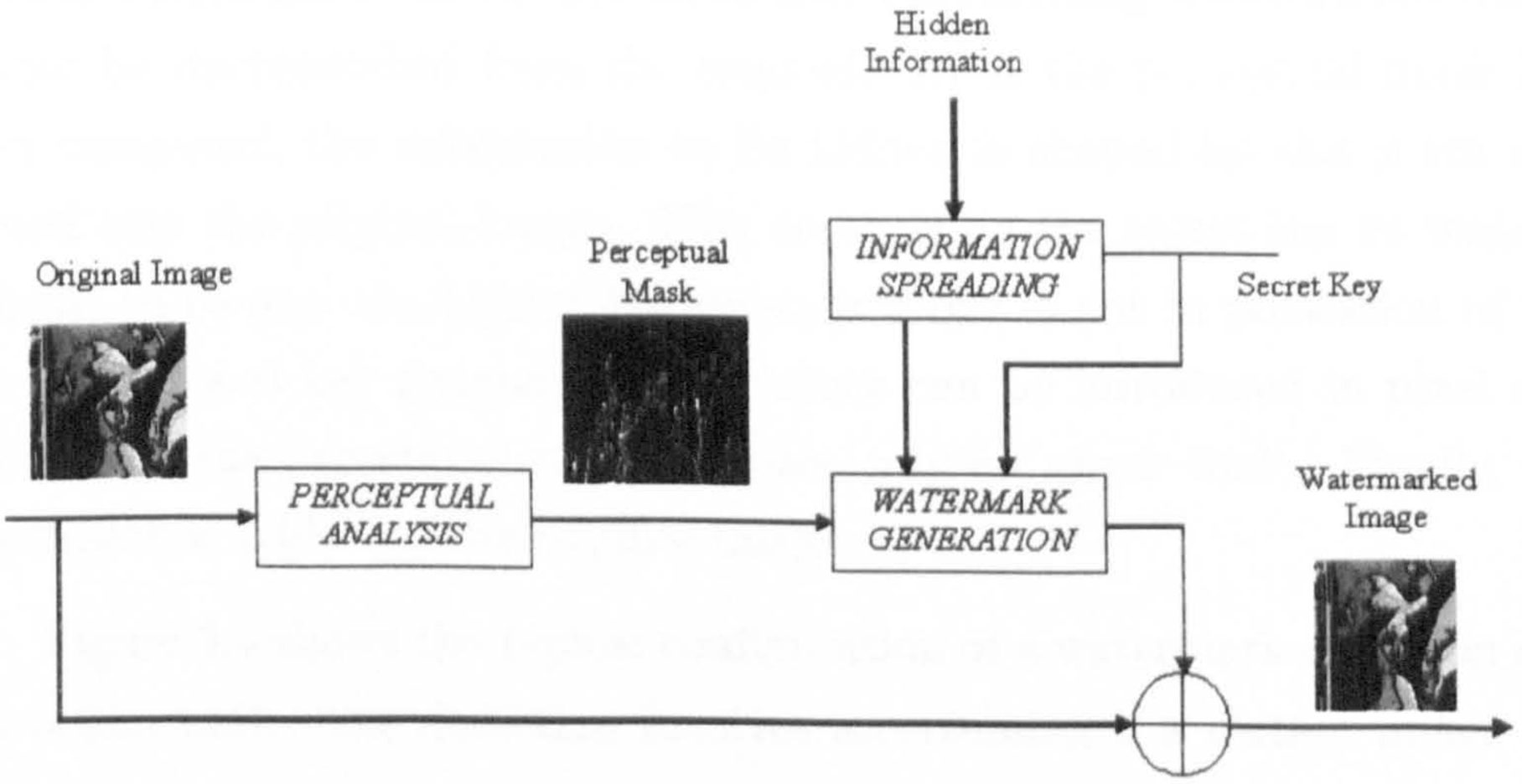


Figure 3.3: Watermark insertion unit

Every watermarking system consists at least of two different parts: watermark embedding unit and watermark detection and extraction unit. Figure 3.3 shows an example of embedding unit for still images. The unmarked

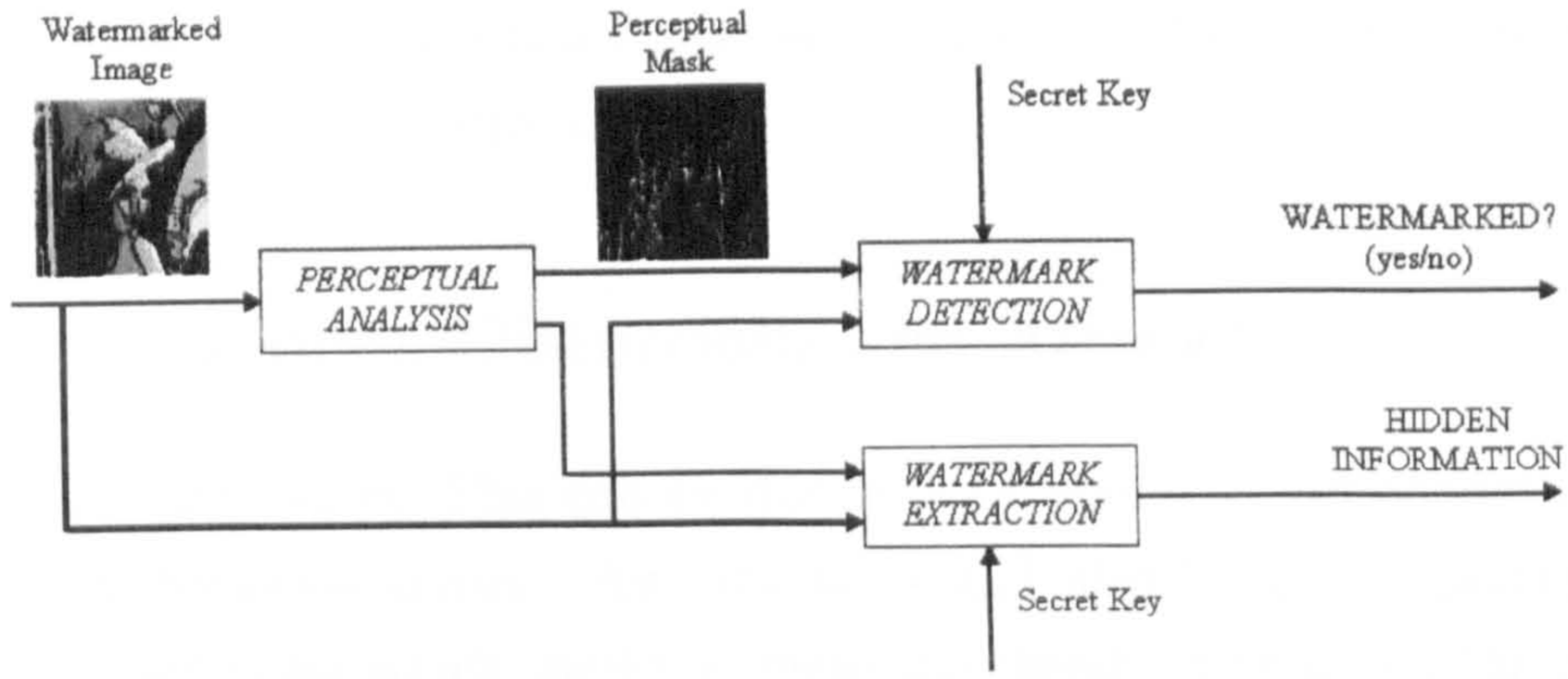


Figure 3.4: Watermark detection and extraction unit

image is passed through a perceptual analysis block that determines how much a certain pixel can be altered so that the resulting watermarked image cannot be distinguished from the original. After the perceptual mask has been computed, the information to be hidden is shaped by this mask and spread over the original image. This depends on the secret key to make it difficult to recover the hidden information if one is not in possession of the key. Additional key dependent uncertainty can be introduced in pixel amplitudes as the perceptual mask imposes only an upper limit. Finally, the watermark is added to the original image.

Figure 3.4 shows the typical configuration of a watermark detection and extraction unit. The detection involves ascertaining if a certain image has been watermarked with a given key. The watermark detector then produces a binary output. Parameters of importance here are the probability of correct detection and the probability of false alarm. These two measures are often used to compare various watermarking schemes[16]. Watermark detection is usually done by correlating the watermarked image with a locally generated version of the watermark at the receiver end. The correlation gives a high value when the watermark has been obtained using the proper key. Once

the presence of the watermark has been correctly detected, it is possible to extract the hidden information.

3.3.1 Common Distortions and Attacks

- Additive Noise: This can be due to D/A and A/D converters or from transmissions errors. An attacker could also introduce perceptually shaped noise which would increase the threshold at which the correlation detector works.

- Filtering: Low pass filtering can affect the performance since spread spectrum like watermarks have a non negligible high frequency spectral contents.

- Cropping: This is when an attacker is interested in a small portion of the watermarked object/image. Therefore inorder to survive the attack, the watermark needs to be spread over the dimensions where this attack takes place.

- Compression: This is common in multimedia applications, e.g. audio, video, images distributed over the Internet. In order to survive this type of attack, the watermark has to be inserted in the same domain as the compression. As an example, DCT domain image watermarking is more robust to JPEG compression than spatial domain watermarking.

- Rotation and Scaling: This type of attack is particularly successful with still images. Correlation based detection and extraction fail when rotation or scaling are performed on the watermarked image because the embedded watermark and the locally generated version do not have a common spatial pattern. Some authors have proposed the use of

Fourier-Mellin transform (rotation and scaling invariant transform) but it makes data hiding difficult.

- Statistical Averaging: This occurs when an attacker estimates the watermark and then removes it by subtracting the watermark from the original image.

- Multiple watermarking: This is when an attacker watermarks an image that is already authentically watermarked. A good approach to tackle it would be to timestamp the watermark by a certification authority.

There are several watermarking methods proposed for various media such as image, audio, video and 3-D geometric models. The early watermarking methods include modifying the least significant bit of the data[17], embedding a secret information that resembles quantization noise[18], and etc. The reader is referred to [19-26] for excellent surveys of the early approaches.

Watermarking schemes achieved more imperceptibility and robustness by transforming the data globally using Fourier transform, Discrete Cosine Transform (DCT) or wavelet transform. To prevent a watermark from being removed without extreme degradation of data, a watermark is embedded into the most perceptually significant portion of the data by analysing frequency domain. Ruanaidh et al[27,28] modified the phase values of the Fourier coefficients to convey the information. Cox et al[25,26] used a spread spectrum method for information embedding to achieve greater robustness. Praun et al[29] addressed robust 3D mesh watermarking by generalising the spread spectrum approach for arbitrary triangular meshes. In some other schemes, the human visual or the auditory systems have been used to generate a more effective watermark. Boney et al[22] generated a watermark by approximating the frequency masking characteristics of the human auditory

system. Podilchuk and Zeng[30] employed visual models to determine image dependent upper bounds on watermark insertion, and increased robustness to common image modification.

In the mid-1990s, the military's interest in unobtrusive communications and public concerns over government efforts to control cryptography, started to drive rapid developments in information hiding[31]. Information hiding enables data to be hidden in other data, for example when a secret message is hidden in an MP3 audio file or a software's serial number is embedded in the order in which certain instructions are executed. Copyright marks do not have to be hidden to be effective. Some TV stations embed their logo in a visible but unobtrusive manner in the corner of the picture and many ECMS systems have control tags bundled quite visibly with the content.

The DVD Marking Concept: The DVD consortium is aiming at finding a copyright marking scheme that will enforce serial copy management. The main idea is that for each disk, a ticket X is chosen which is a random number in addition to copy control information and some data unique to the physical medium. A one-way hash function h is used to computer h(X) and h(h(X)). This is embedded in the video as a hidden copyright mark. Machines compliant with this should look for a watermark and refuse to play the track unless supplied with h(X). This is checked by hashing it and comparing with the mark. In other words a compliant device will record a track only when supplied with X, and h(X) is written to the new disk. Thus a 'copy once only' track in the original medium becomes a 'copy no more' track in the new medium[31].

Copy generation management using embedded marks, rather than attached data is quite robust for digital to analogue conversion and vice versa. The disadvantage however is finding a method of embedding a mark in audio or video which might be difficult to embed but can be detected readily and

difficult or impossible for an attacker to remove. Detection must be done by cheap mass market equipment with a low false positive alarm rate.

3.3.2 Attacks on Copyright-Marking Schemes

- In the beginning, many people assumed that the main market would be watermarking- embedding hidden copyright messages so that the ownership of a work could be proved in court. This has turned out to be untrue. Intellectual property lawyers do not seem to have a problem in proving the ownership of an article and they do not rely on technical measures which may confuse the jury, instead they rely on documents such as contracts with bands and model release forms. The legal use of copyright marks may rather be for fingerprints like hidden serial numbers.

- Digimarc set up a service to track intellectual property on the Web. This system proved to be useful in that one of the main costs faced by multimedia producers is tracking the copyright of large numbers of images and the royalties due to their owners. However, it was found that the Digimarc system could be easily defeated by guessing the master password or modifying the marking software so that it would overwrite existing marks. Digimarc developed the 'Marc Spider', a bot that crawled the Web looking for marked pictures and reporting them to the copyright owners. Again it was found that there are several ways to defeat this. More details on Digimarc products have been mentioned later in this chapter.

- Many marks are said to be additive. This can cause a lot of problems, for example, say if all the frames in a video carry the same mark, it is possible to average them to get the mark and then subtract it out.

An even easier method is to supply some known content to a marking system, and then compare its input and output. Consider, there is a picture P, with a mark m. We have here a system P+m. A competitor whose mark perhaps is m′ might say that the original was P+m+m′. This means that the published picture P+m was infact marked with m′.

- According to the Kerckhoffs' principle, security of a system should reside in the choice of key, not in the algorithm in use. But many designers simply ignore this principle. This principle becomes useful when marks are to be used in evidence as they have to be revealed in court. Infact, as even the marking keys will have to be disclosed, it may be necessary to protect objects with multiple marks.

- A lot of research has been done to develop a marking equivalent of public key cryptography[31]. For example, inserting a mark that only one principal could detect or anyone could detect a mark that only one principal could have inserted. The first option is possible if the mark can be inserted as the cover audio or video is being manufactured. Given a device that will detect the mark, an attacker can remove a mark by applying small changes to the image until the decoder cannot find it anymore.

- More steganalysis techniques were developed to break particular embedding schemes. As an example, when the mark was added by either increasing or decreasing the luminosity of the image by a small fixed amount, this caused the peaks in the luminosity graph to become twin peaks, which meant that the mark could be filtered out over much of many images.

- A typical Web browser, when presented with a series of graphics, will

display them continuously without any gaps, so a marked image can often be broken down into smaller images which together will look exactly like the original when displayed on a Web page, but in which a copyright mark will not be detected.

- The most general known attacks on copyright marking schemes involve suitably chosen distortions. Audio marks can be removed by randomly duplicating or deleting sound samples to introduce inaudible jitter; techniques used for click removal and resampling are also powerful mark removers. A tool called Stirmark[31] exists which introduces the same kind of errors into an image as printing it on a high quality printer and then scanning it again with a high quality scanner. The image is slightly distorted geometrically- it is slightly stretched, sheared or shifted and/or rotated by an unnoticeable random amount. This system defected all the then existing marking schemes and is now a standard benchmark for copyright mark robustness. The idea is to design marking schemes that will resist a chosen distortion attack, in which the attacker who understands the marking scheme distorts the content in such a way as to cause maximum damage to the mark and minimal damage to the marked content.

A more detailed account of copyright marking schemes can be obtained from [32-33].

The technology is improving consistently but the drawback appears to be designing marking schemes that remain robust once the mark detection algorithm is known.

3.4 General Information-Hiding Techniques

Until recently, hiding confidential information was considered much more important that enciphering it. Military organisations still believe in this view and use techniques such as microdots (used by spies) to the low-probability-of-intercept radios.

When it involves hiding data in other data, the copyright mark or in the case of steganography, the embedded text, is hidden in the cover-text producing the marked text or in steganography it is called stego text. Some additional secret information is used during this process called the marking or the stego key and is needed to recover the mark or the embedded text. Text can be replaced by audio, video etc.

There are a number of embedding schemes proposed. To mention a few:

- The obvious technique is to hide the mark or the secret message in the least significant bits of the audio or the view signal. This is commonly used in many public domain steganography tools. Its main drawback is that the hidden data is easy to detect statistically as the least significant bits are no longer correlated with the rest of the image and is trivial to remove or replace. It can also be damaged due to lossy compression techniques.

- Another method first invented in classical China is to hide the secret message in a location determined by the secret key[31]. In this case, both the sender and the receiver had copies of a paper mask which had holes cut out of it at random locations. The sender would place his mask over a blank sheet of paper, write his message in the holes, then remove it and compose a cover message including the characters of the secret embedded message.

- A more recent adaptation of the above technique hides a message in a .gif format[31]. The secret key is expanded into a keystream, which selects an appropriate number of pixels. The embedded message is the parity of the colour codes for these pixels. In practice, even a large number of the pixels in an image can have their colour changed to that of a similar one in the palette without any visible effects. If all the pixels are tweaked in this way, then the hidden data is easy to remove by just tweaking them again. A better result is obtained if the cover image and embedding method are such that say only 10% of the pixels can safely be tweaked. Then if the process is repeated, but with a different key, an independent 10% of the pixels will be tweaked and only 10% of the bits of the hidden data will be corrupted.

- The introduction of noise or distortion will introduce errors in the hidden data irrespective of the embedding method unless error correcting code is added. The system used for banknote marking called Patchwork, uses a repetition code - the key selects two subsets of pixels, one of which is marked by increasing the luminosity and the other by decreasing it. This embeds a single bit; the note is either watermarked using that key or it isn't. Most common techniques employed are the direct sequence spread spectrum techniques borrowed from electronic warfare.

- Often spread spectrum encoding[30] is done in a transform space to make its effects less perceptible and more robust against common forms of compression. These techniques are used along with perceptual filtering, which emphasizes the encoding in the noisiest or perceptually most significant parts of the image or music track, where it will be least obtrusive and de-emphasizes it in quiet passages of music or large expanses of colour.

- Another method for marking print media exists which works by moving text lines up or down by a three-hundredth of an inch, or adding extra echoes to music below the threshold of perception. Such techniques are not yet robust as generic ones based on keyed embedding using transform spaces, spread spectrum and perceptual filtering.

Watermarking can also be divided into two operational categories: private watermarking where the public is not allowed to access the watermarks in any way and public watermarking where the public is only allowed to detect the watermarks and nothing else[34]. As an example, consider a corporate website containing special images. Each original image is accompanied by a number of versions enhanced by some image processing tools. Each of the enhanced images is watermarked with a text string describing the process that has been applied. Given the original and the enhanced images, an informed detector reveals the string describing the enhancement process. Therefore, here we have an informed detector being used in a public watermarking application.

A lot of applications require watermarks to be detected in works that may have been altered after embedding. Robust watermarks are designed to survive normal processing but secure watermarks should resist any attempt to thwart their purpose. In most applications a watermark cannot perform its function if it is undetectable, therefore robustness is an essential property if a watermark is to be secure. However, robustness alone cannot be sufficient for security applications as secure watermarks must be capable of surviving novel processes, for example lossy compression, digital-to-analog-to-digital conversion, analog recording such as VHS or audio tape, printing and scanning, audio playback and re-recording, noise reduction, format conversion etc. that are designed to remove them.

In practice, watermarking systems employ several strategies for han-

dling various types of distortion. Image watermarking systems commonly use redundant embedding across several coefficients. If some of these coefficients in the image are damaged, the watermark in other coefficients should remain detectable. For more information on using this method and ways of recovery of the watermark, please refer to [34]. It is not necessary that the computation of watermark embedding be performed in the Fourier domain. A low frequency reference pattern can be easily represented and embedded in other domains such as the spatial or temporal domains. Basically the description and analysis of a watermark can be performed in one domain but the embedding and detection can be done in any other domain.

Another method that has been used in the frequency domain to embed watermarks is the spread spectrum coding method. In this system, messages are encoded with sequences of symbols which are transmitted in a temporal sequence, each one being represented by a signal called as chip. Chips are pseudo random sequences of 1s and 0s spread across a wide range of frequencies in the frequency domain. This means that if the signal is distorted by a process that damages only a fraction of the frequencies, such as a band pass filter or addition of band limited noise, then the chips will still be identifiable. Here the two main characteristics important to watermarking are the low signal to noise ratio reduces the risk of perceptible artifacts. However, the detector output signal may still have a high signal to noise ratio as it despreads or concentrates the energy present in a large number of frequencies. Also. The watermark dispersed over a large number of frequencies makes it robust to many common signal distortions.

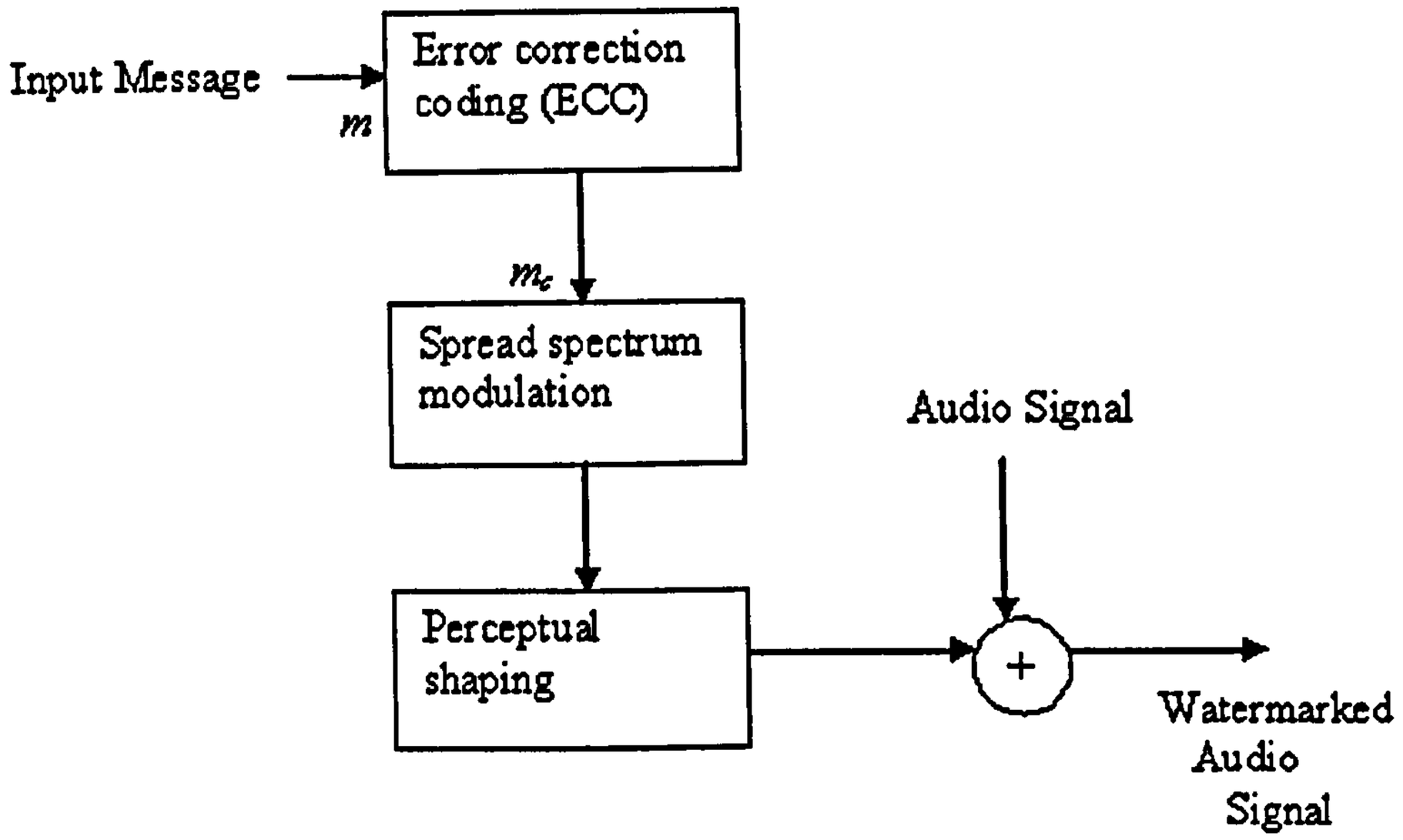


Figure 3.5: Simple Spread Spectrum audio watermarking method

Another method commonly used is embedding the watermark in perceptually significant coefficients. Some coefficients are so susceptible to distortions that they are unlikely to be useful. Most image processing techniques damage the high frequencies in the Fourier representation of an image. Lossy compression quantizes many high frequencies to zero, half-toning adds high frequency noise, etc. If the information in such unreliable coefficients is used during detection, it will reduce the reliability of the whole system. For further information please refer to [34].

Yet another method is embedding the watermark in coefficients of known robustness. As an example, if the image watermark has to withstand spatial shifting and linear filtering, then the watermark might be embedded in the Fourier magnitudes of the image. Other domains such as the log-polar Fourier Transform or the cepstrum transform can also be used. Identifying the coefficients that best survive the distortions can be done either analyt-

ically or by empirical tests, which involves comparing the content directly after embedding and before detection. It is obvious that such experiments will have to be performed over a wide variety of content and numerous trials are needed to build a satisfactory model with sufficient statistical reliability[34].

Most methods of watermark embedding attempt to create a watermark that remains relatively unchanged after normal signal processing. A relatively different approach is to attempt, during the detection process, to invert any processing that has been applied since the watermark was embedded[34]. As an example, a clockwise rotation of an image can be inverted by a counter clockwise rotation of the same angle. Obviously, for rotations that are not multiples of 90°, this inversion is likely to approximate due to interpolation and round-off error.

It is sometimes possible to explicitly identify parts of the image etc. that are robust to expected attacks and embed the watermark in these. If the detector used is informed, this can even be done on a per image basis. If the detector can determine the process that has been applied to the image, then it can invert the process or apply it to the reference mark.

Research is being done on increasing the robustness of watermarkes to geometric distortions like temporal delay or spatial scaling as they are more difficult to handle than valumetric distortions. There are a few approaches proposed by Cox and Bloom[34] to this problem. Exhaustive search entails inverting a large number of possible distortions and testing for a watermark after each one. As the number of possible distortions increases, the computational cost and false positive probability can become unacceptable. However, synchronisation and registration patterns can be embedded to simplify the search, thus preventing an increase in the false alarm rate and also computationally more efficient. But the registration pattern may not be detected

and also the watermark may be undetected after registration. A type of watermarks called invariant watermarks can be constructed using the log-polar Fourier transforms. These can survive some forms of geometric distortions and so are quite robust.

In some applications, even the minimal distortion introduced by embedding a 160-bit signature might be unacceptable. A classic example is in medicine where any modification of an image can be considered as malpractice. This has lead to the concept of erasable watermarks. The method employed is as follows: A cryptographic signature is embedded in the article to be watermarked in an erasable manner. The person who receives it extracts and records the embedded signature. The watermark is then erased at which point the article becomes exactly identical to the original unwatermarked work. For verification purposes, the recipient computes a one way hash of the article and compares it with the hash decoded from the sender's signature. If both the computed and the received hashes are identical then the article is authentic. However, the fundamental problem this concept raises is that in theory, it is impossible to make an erasable watermark that can be embedded in 100% of digital content and also has a high false positive probability.

Semi-fragile Watermarking by Quantizing DCT Coefficients: An image authentication system called the E_DCTQ/D_DCTQ uses a semi-fragile watermark designed to survive specific levels of JPEG compression[34]. In the JPEG image compression system, images are quantized in the block DCT domain where the quantization step size for each coefficient depends on its frequency. By default, the step sizes are obtained by multiplying a predefined quantization array by a given constant factor (Table 3.1).

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Table 3.1: Luminance quantization matrix used in JPEG. The upper left value (16) is the base quantization factor for the DC term of each 8x8 block. The lower right value (99) is the base quantization factor for the highest frequency term. These base values are multiplied by a global quantization value to obtain the actual quantization factor used[34].

Table 3.1 shows the quantization array used. The constant factor is usually a simple function of a "quality factor" between 0 and 100, which is entered by the user. Although it is possible for JPEG encoders to use more sophisticated algorithms, e.g. designing a unique quantization array for each image. However, the authentication system here is targeted at the default JPEG behaviour.

Fragile and semi-fragile watermarks are often not secure against malicious tampering as they succumb to copying. Also, semi-fragile watermarks are only able to authenticate those properties of an image they are embedded within. For example, in the watermarking system mentioned above (E_DCTQ/D_DCTQ) the authentication watermark is only embedded in the high frequency coefficients of the block DCT, as it becomes too visible if embedded in the low frequency coefficients. This makes it obvious that only changes in the high frequency will affect the watermark and any changes in

the low frequency with the high frequency remaining unaffected, will report the system as being authentic.

There are two main advantages of using a system of this type. The first approach being that the counterfeiter cannot make a copy by one simple attack as each image has a different watermark embedded. And secondly, the signature is based on properties of the image that cannot be changed without causing unacceptable fidelity problems.

Another type of watermark called the tell-tale watermark[34] exists which means as the name suggests, it is a way of investigating how an image is corrupted by examining how the known embedded watermark has been affected. This type of watermark would distinguish between a wide variety of distortions. Research in these schemes is still in the early stages however, substantial amount of research is being done into the problem of using watermarks to determine where an image has been affected.

The types of security required of a watermarking system depends on the application. Applications may require the prevention of one or more of the following scenarios:

1. Unauthorised embedding
2. Unauthorised detection
3. Unauthorised removal
4. System attacks

Each of the above type of restrictions involve varying degrees of security.

A new digital image watermarking system that complements a wavelet based insertion module, with a resynchronisation module and a method for selecting the watermark using an estimated quality based average has been

proposed by Fullea and Martinez[35]. This technique seems quite robust and has been tested intensively with attacks performed by Stirmark with results of robustness over 90%. The proposed systems comprises of three modules: Insertion module which inserts the watermark in the digital image; Resynchronisation module which estimates a possible transform caused by geometric attack and reverses it prior to the watermark extraction process. This makes the system robust against signal processing attacks like quantification, compression, filtering etc. and geometric ones like rotation, shift, scaling etc. The extraction module recovers the watermark from the resynchronised distributed copy making use of the original image. The output of this module is coupled to an additional error correction module which depending on the application will present a specific reliability/number of watermarks rate which means a specific rate between the number of bits used for error correction and identification. The advantage of the technique is that the system parameters are configurable and therefore can be adapted to the particular application. For further information please refer to [35].

Another watermarking scheme proposed by Erard, Kutter, et al[36], provides data hiding and retrieval with the use of a secret key. The unmarked image is not required to retrieve the hidden information and an integrity check on the embedded data can also be performed. The secret key generates the binary random sequence and it is mixed with the actual watermark before the embedding process. Thus, without the knowledge of the key, it is impossible to extract the watermark, let alone prove its existence in the image. The CRC-16 checksum method is used to assert the integrity of the retrieved watermark. This technique is based on the amplitude modulation technique[37].

Data integrity verification is extremely important to judge the authenticity of the data as it is quite easy to tamper with digitised data and leave

no traces. Verification of data has been broadly classified into two main categories: digital signature based and watermark based. The difference between the two being digital signature is a representation of the characteristic of the image, stored as an extra file and used at a later stage for authentication. Alternatively, watermarking embeds secret information in an image which is extracted using various methods to verify the authenticity of the data. A good authentication scheme should be robust to tampering and attacks. According to the underlying technology used, the above methods can be classified into quantization based[38-41], feature based[42-43] and relation based[44-45]. Kundur and Hatzinakos[38] designed a wavelet based quantization process which has a disadvantage in that the method cannot resist incidental modifications and the tampering detection results are very unstable. Mark Liao, Lu et al[46] have proposed a multipurpose watermarking scheme for image/audio authentication and protection. Here they combine a host data dependent quantization and a complementary watermark hiding strategy[47] to conceal watermarks and present various detection methods with a high degree of robustness and fragility simultaneously.

For feature based authentication system, Bhattacharjee and Kutter[42] proposed to generate a digital signature for encrypting the feature points of an image. The positions of the feature points are then compared with those decrypted from the previously encrypted feature points. The disadvantage here is that it is not shown if the approach can resist JPEG compression with middle-to-high ratios. Dittmann et al[43] presented a content based digital signature approach for image/video authentication using edge characteristics. This is similar to [42], but different extraction techniques are used.

In order to make the designed image authentication system survive JPEG compression, Lin and Chang[44-45] proposed to preserve the invariance between the DCT coefficients before and after quantization. However,

it does not show how the method can withstand other incidental manipulations. Here although the invariance property is used to authenticate images, it should be noted that this relationship is random because the invariance property of any two 'random' DCT blocks are stored as the digital signature.

Liao and Lu[46] have proposed a new image authentication scheme where the 'structure' of an image's characteristics is considered as the digital signature. The structure of the image content is composed of parent-child pairs in the wavelet domain. The method is robust to content-preserving manipulations and fragile to content changing distortions.

Researchers at the Purdue University, US have proposed a new 2-D watermarking scheme called the Variable-Watermark Two-Dimensional Algorithm (VW2D)[50-51]. It involves the use of semi-fragile, spatial watermark and the technique generates a pseudo-random sequence from a small key file, which is uniquely associated with an owner. The sequence, consisting of either 0,1 or -1,1, fills multiple 8x8 pixel blocks to eventually cover the entire original image. The collection of blocks forms the full watermark, which is the size of the original image. The marked image is formed by arithmetically adding the watermark to the original pixel data.

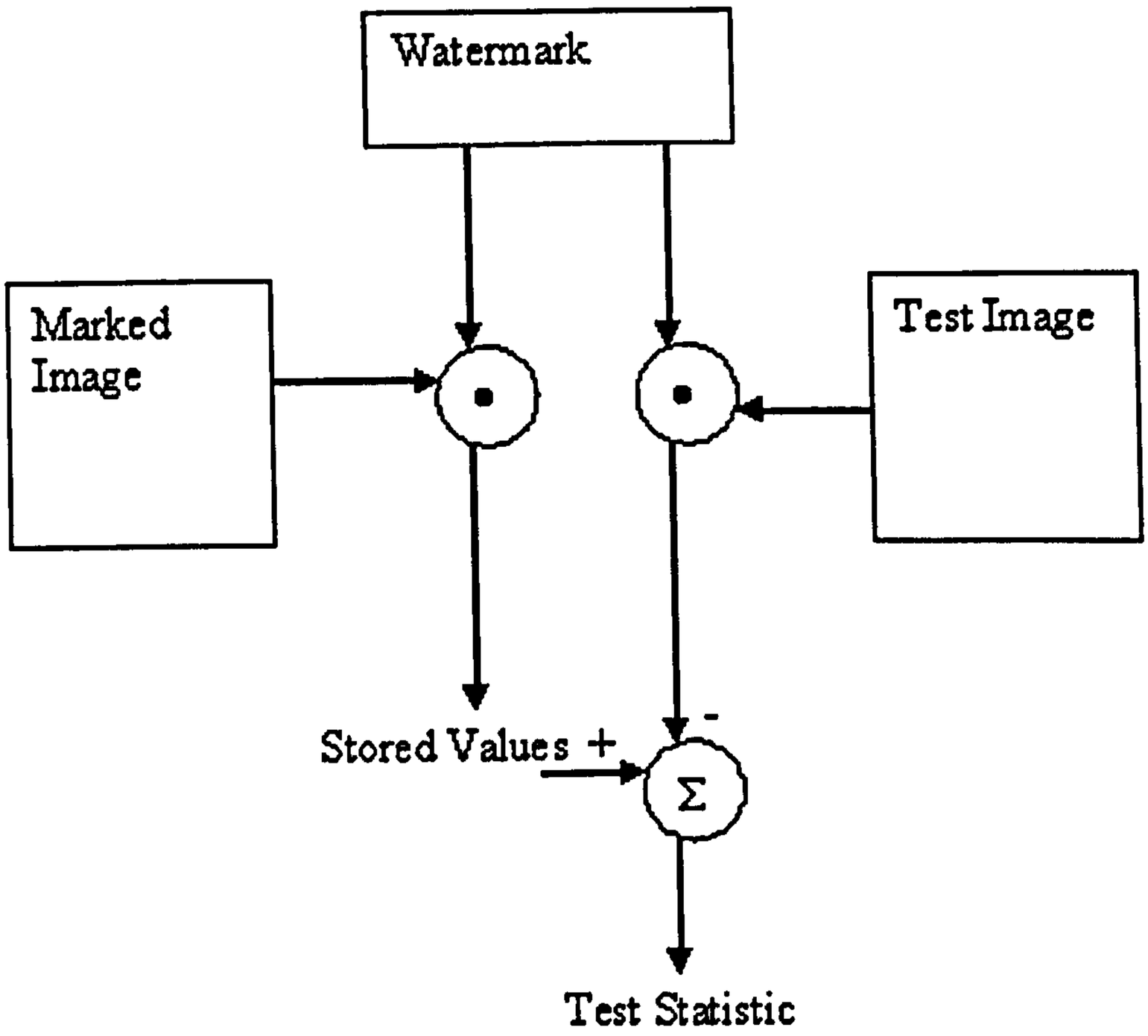


Figure 3.6: VW2D testing procedure

With reference to the above testing procedure schematic, as part of the marking procedure, the inner (dot) product between each watermark block and corresponding marked image block is computed. Testing consists of computing the inner product between a test image block and the mark and subtracting this from the previously stored value. This algorithm has been shown to detect linear and non-linear filtering and the JPEG compression.

3.4.1 Image adaptive watermarks in the DCT and Wavelet Domain

Research on robust, image-adaptive (IA) watermarks is being conducted at Bell Labs, Lucent Technologies[30]. The image adaptive watermarks are based on the robustness of the basic spread spectrum technique. This technique embeds a normally distributed zero mean pseudorandom sequence, into the transform coefficients (DCT or wavelet) of an original luminance image. To verify a possibly altered marked image, the normalized correlation coefficient (test statistic) between the watermark and the version of the watermark extracted from the marked image is obtained. A value near 1 indicates the presence of the mark; a small value (<0.1) indicates that the watermark under test is not in the image, or that the image has been severely and visibly degraded to the point of removing the watermark. Figure 3.7 shows the Image Adaptive watermarking procedure. Another watermarking algorithm proposed at Purdue University can be described by the following two figures. Here Delta is a measure of how much the watermark has been altered as a result of changes to the images. The results have shown that the watermarks detect all but the most minute changes to an image. A two dimensional Digital Watermarking method has been proposed by Tirkel, Schyndel, Osborne[50], where a robust, undetectable, digital watermark has been coded on a standard 512*512 intensity image with an 8 bit gray scale. This watermark is capable of carrying authentication or authorisation codes, or a legend required for image interpretation thus finding applications in image tagging, copyright enforcement, counterfeit protection and controlled access. The method is based on linear addition of the watermark to the image and Least Significant Bit (LSB) manipulation followed by correlative recovery. The watermarks here are chosen from two-dimensional array patterns based on m-sequences[51,17] or extended m-sequences.

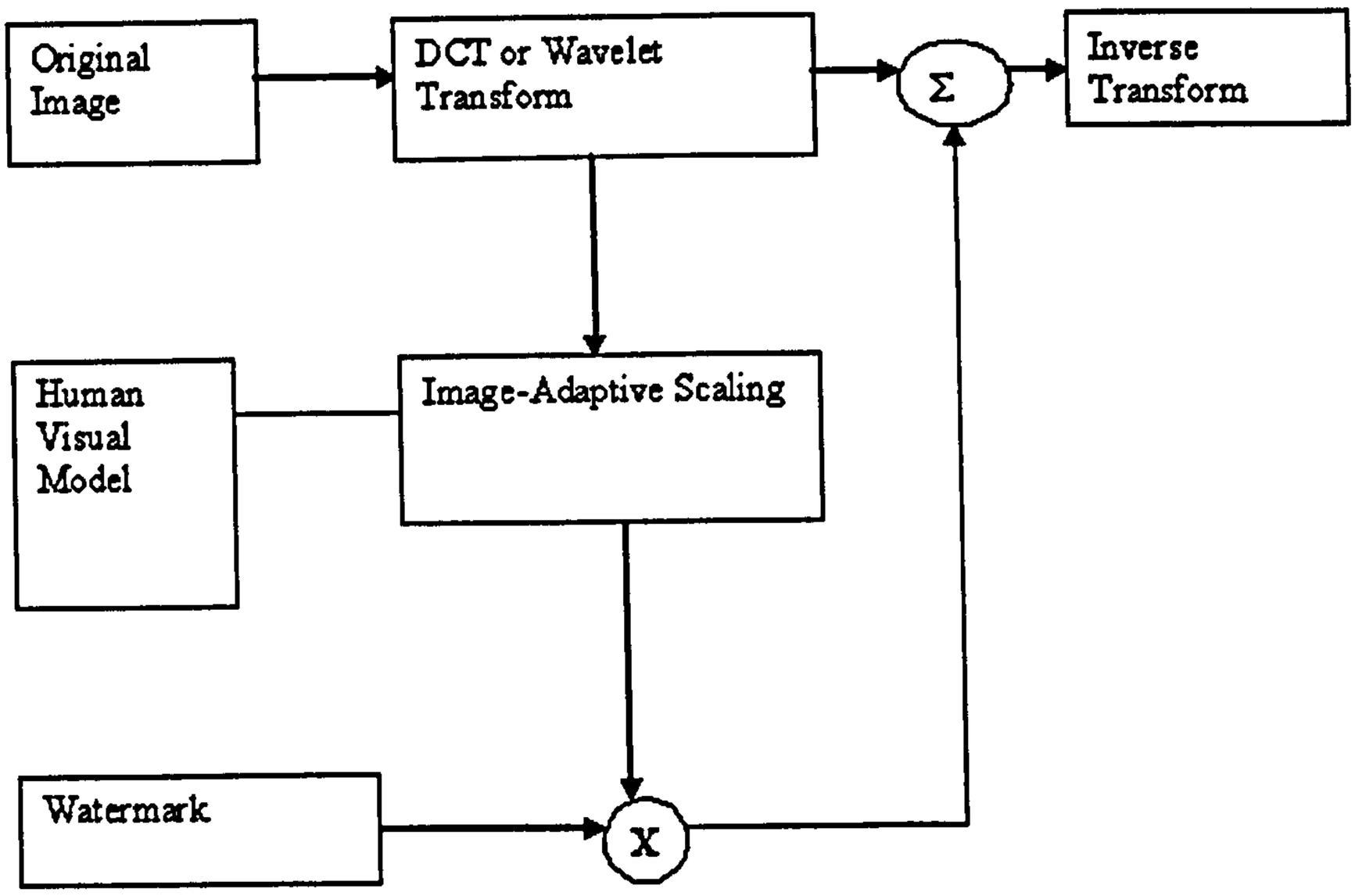


Figure 3.7: Image-adaptive watermarking procedure

Walton[52] in 1995, proposed a novel technique involving a fragile watermark, where, by deliberate design, any distortions render the watermark non-recoverable and this serves as proof of tampering. The method also uses LSB manipulation and introduces as effective palette manipulation technique to increase the watermark effectiveness by involving complete RGB image components.

The FRFT (Fractional Fourier Transformation) domain watermarking concept has been proposed by Djurovic, Stankovic, Pitas in December 2000 [53]. This method results in the possibility of generating more watermarks in the FT and DCT domains. Results here show that the method is more robust on many attacks performed by pirates.

Su and Kuo[54-56] proposed three watermarking schemes consisting of complete watermark embedding and detection system. The first method

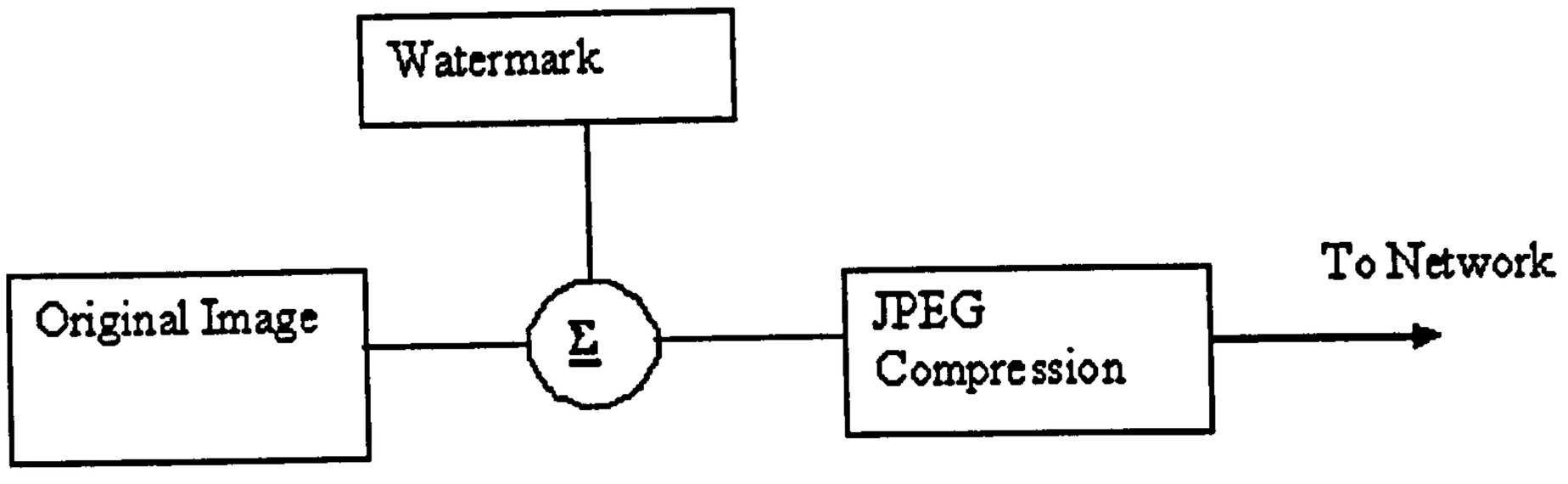


Figure 3.8: Image Watermarking Algorithm

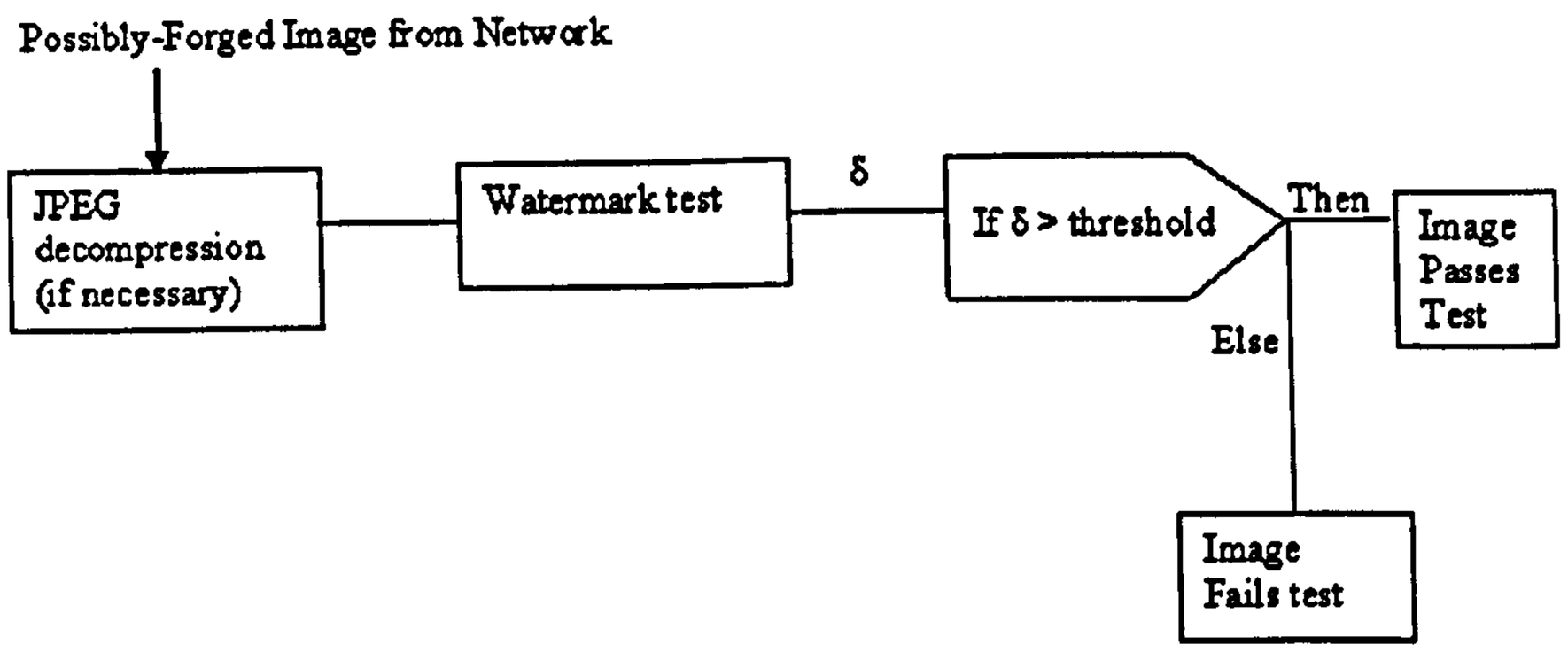


Figure 3.9: Watermarking Testing Algorithm

is an integrated approach to image compression and watermarking. This coding scheme is EBCOT (Embedded Block Coding with Optimized Truncation), accepted by the JPEG2000 compression standard as the core coding algorithm. Due to this, the embedding and retrieval system is more efficient than the existing schemes. Both the embedding and detection occurs in the compressed domain to eliminate the need to compress the host image after embedding and decompress the watermarked image before detection. The embedded watermark is robust against various signal processing attacks including compression and filtering while the resulting watermarked image

maintains good perceptual quality.It can be detected progressively without waiting for the whole image to be downloaded. The embedded watermark can be detected without the reference of the original image so that it is a blind watermarking scheme.

Most existing watermarking schemes are based on the additive spread-spectrum method because of its robustness against noise and distortions. However, they fail to detect the watermark when the image undergoes geometrical modifications such as cropping, rotation, scaling, etc. This is due to the loss of synchronization between the watermark detector and the embedded watermark. Since the possibly existing watermark is hidden in a very strong noise, i.e. the image content, it is very difficult for the watermark detector to predict the correct position of the hidden watermark when the image is cropped, scaled or rotated. Cropping or scaling can cause information loss of the watermark so that the embedded information may not be correctly determined.

To overcome the above drawback, a spatial-frequency composite watermarking scheme has been proposed. Two watermarks will be embedded into a digital image using the spread spectrum approach, one in the spatial domain and the other in the frequency domain. The frequency-domain watermark contains the desired hidden information that will be carried with the host image. The spatial-domain watermarking method is used to achieve self-registration of the investigated image with the original image. After the registration process as a result of spatial-domain watermark detection, the hidden information can be successfully determined from the registered image by detecting the frequency-domain watermark. Both spatial- and frequency-domain watermark detection processes are blind. The watermark can survive a combination of manipulations, including cropping, scaling, rotation, shearing, change of the aspect ratio, column/row dropping and linear trans-

formation.

Block-based watermarking is a general type of scheme, in which the image is divided into blocks for watermark embedding and detection. With divided blocks and their local statistics, the balance of robustness and invisibility of the watermark can be achieved. Block-based Discrete Cosine Transform (DCT) is utilized in many existing watermarking schemes. By decomposing the image into several frequency bands using DCT, the watermark is embedded into the significant frequency coefficients to attain a robust watermark. The watermarked image is then formed by applying inverse DCT on each image block. The sensitivity of the human visual system to DCT basis functions has thus been extensively studied. It is advantageous to use the visual model within this framework in block-based watermarking systems to reduce the impact of quality degradation and, at the same time, to make the watermark survive JPEG compression attack better. Moreover, spatial division can be used to increase the amount of the embedded watermark, i.e. groups of blocks are embedded with different information so that multiple bits can be carried by a single image. Security of the system can also be increased by scrambling the order of the blocks so that the watermark cannot be detected correctly without the correct descrambling process.

Existing researches on block-based watermarking focus on the robustness against filtering/compression attacks and perceptual issues. However, geometrical attacks are very difficult to resist, especially for the block-based methods. The geometrical attack will seriously limit the usage and applications of the related researches on block-based watermarking.

Watermarking schemes based on the Discrete Wavelet Transform(DWT) are also becoming increasingly used, with the development of the JPEG2000 standard. Research in this area is in the initial stages, though evidence exists that DWT could enhance the robustness of the watermark against

attacks[57]. Mandyam and Ahmed introduced the DLT (Discrete Laguerre Transform) in 1996[58]. It has been shown by Gilani and Skodras[59] that the quality of the DLT domain watermarked images is higher than the corresponding DCT domain watermarked images, though from the robustness viewpoint, both have similar performance.

A wavelet based reversible watermarking system proposed by Jun Tian from Digimarc Corporation[60] can be applied to audio and video. This method uses an integer wavelet transform to losslessly remove redundancy in a digital image to allocate space for watermark embedding. The algorithm starts with a reversible colour conversion transform, then the integer wavelet transform is applied to one or more decorrelated component/s. These two transforms remove the irregular redundancy in the digital image so that regular redundancy can be embedded in the form of hash of the image, compressed bit stream or any other image content dependent watermark. An SHA-256 hash of the original image is embedded for authentication purposes.

An Adaptive Digital Image Watermarking Technique for Copyright Protection has been proposed at the University of California, Santa Barbara[61]. Here the watermark embedding consists of two steps: the scrambling of the watermark image and the insertion into the host image. The recovery is then the reverse of these two steps.

In the scrambling stage, the pixels of the watermark, each with only one bit of data, are pseudo-randomly permuted to form a new watermark image. This permutation is done using a linear feedback shift register. By setting the state of the shift register, a pseudo random sequence can be generated that is recoverable by resetting the shift register to the original state. There are two ways to use the shift register, one is to use it to generate a random sequnece of new row and column induces for the two dimensional watermark. The second, which is easier to implement involves performing a raster scan

of the watermark to generate a single row vector from the watermark. The elements of this row vector are then pseudo-randomly permuted into a new vector via a single execution cycle of the linear shift register. The shift register can only perform one permutation of the indices of the raster scan vector. A new raster scan vector is then generated by assigning the elements from the old one to the positions of the new vector. An inverse raster scan process is performed on this vector to produce the scrambled watermark.

Once the binary watermark is scrambled, it is inserted in the image. This is done by inserting the pixels individually into blocks of pixels of the host image in a pseudo random method. The image to be watermarked (host image) is divided into nxn blocks into which one bit of the watermark is embedded. This depends on the size of the host image and the number of pixels that make up the watermark.

This algorithm has been shown to be robust to low pass filtering, median filtering, scaling, cropping, rotation and lossy JPEG compression[61]. An embedded signature image can be recoverable and recognisable even after the watermarked image has been tampered with by image processing techniques.

3.5 Applications of Digimarc Technology

Digimarc ID Systems designs secure ID documents that combine a wide range of security features and technologies into a layered card that meets customer security requirements within their budgets. Their ID cards can use a range of security features including microprinting, optically variable devices, digital watermarking and multicolor UV printing in layered identity cards. The increasing use of digital imaging systems is making it easier than ever to duplicate, alter and distribute images. In a security environment, it is becoming increasingly vital to manage and track digital images securely.

Digimarc ID Systems card technologies are used in a variety of applications including driver licenses, child identification, national identification, voter identification and smart cards. It is an integration of the technologies to create complete, cost-effective and state-of-the art identification solutions.

3.5.1 Defence and Intelligent Imaging

Digimarc digital watermarking as a security feature in digital images enables:

- Management and distribution of digital images
- Tracks content as it travels
- Adds a covert digital identity that stays with the content as it is distributed

3.5.2 Digital Solutions

Digital systems electronically capture and combine portraits, signatures, and alphanumeric data with any of several preprogrammed, full-color card designs - including such custom components as logos, seals and officer signatures. The completed cards can be previewed prior to printing or saving and can be produced instantaneously using PVC card printers or at a central location on proprietary high-volume card production systems. Digital systems allow data to be stored electronically providing the ability to:

- Update/reissue cards without the subject present or even via the Internet
- Retrieve and transmit an individual's dossier in minutes

- Incorporate biometric security features including finger imaging and facial recognition

3.5.3 Hybrid Digital/Analog Solutions

Sometimes the right solution for an identification system is neither purely digital nor completely photographic, but rather a hybrid of the two.

Hybrid solutions are often a cost-effective means of capturing data in remote locations, where volumes are too low to warrant a full-fledged digital system or the infrastructure is inadequate to support an electronic system.

3.5.4 Driver Licenses

Digimarc ID Systems is the leading supplier of driver license systems in the U.S. They also deliver driver license systems internationally and provides solutions to the United Kingdom, Russia, several Canadian provinces, Costa Rica, Botswana, etc.

3.5.5 National ID

Many countries issue national ID cards in an effort to provide a standardized identity document, reduce illegal immigration and prevent benefits fraud by requiring citizens to present identification for access to government programs and services. Digimarc ID Systems' card materials are durable, tamper-resistant and include security features that prevent counterfeiting or alteration.

3.5.6 Voter ID

Voter ID cards are frequently instituted to ensure the integrity of a country's election process. To help prevent voter fraud, cards can be produced using a secure process that includes security features and even records biometric data.

3.6 Product Capabilities

A Digimarc digital watermark is imperceptible to the human eye. It is a special message embedded in an image i.e. video, photo or other digital content by making subtle changes to the data of the original digital content. The changes caused by watermark embedding can be calculated by subtracting the original image from the watermarked image. Digimarc digital watermarking product capabilities are an effective defense in security applications:

- Applicable to packaging, value documents, plastic cards, holograms at low production cost and without need for additional materials
- Applicable to digital images
- Data carrying
- Covert
- Machine readable by standard reader technology, such as Web cameras and scanners, and our special software
- Compatible with other security features such as holograms, bar codes, special inks and papers

- Resistant to alteration or duplication
- Recognizable under automatic inspection
- Cost effective - low production costs
- Easy to implement
- No design change necessary

However, the actual algorithms being confidential to Digimarc, are not available for testing purposes.

3.7 The Technology

3.7.1 Digital watermarks and methods for security documents

Security documents (e.g. passports, currency, event tickets, and the like) are encoded to convey machine-readable multi-bit binary information (e.g. a digital watermark), usually in a manner not alerting human viewers that such information is present. The documents can be provided with overt or subliminal calibration patterns. When a document incorporating such a pattern is scanned (e.g. by a photocopier), the pattern facilitates detection of the encoded information notwithstanding possible scaling or rotation of the scan data. The calibration pattern can serve as a carrier for the watermark information, or the watermark can be encoded independently. In one embodiment, the watermark and the calibration pattern are formed on the document by an intaglio process, with or without ink. A photocopier responsive to such markings can take predetermined action if reproduction of a

security document is attempted. A passport processing station responsive to such markings can use the decoded binary data to access a database having information concerning the passport holder. Some such apparatuses detect both the watermark data and the presence of a visible structure characteristic of a security document (e.g. a printed seal of the document's issuer).

3.7.2 Multiple watermarking techniques for documents and other data

Multiple digital watermarks, each of which has different characteristics, are embedded in a document. The characteristics of the various watermarks are chosen so that each of the watermarks will be affected in a different manner if the document is subsequently copied and reproduced. The detection process or mechanism reads each of the watermark and compares their characteristics. While wear and handling may change the characteristics of the digital watermarks in a document, the relationship between the characteristic of multiple digital watermarks will give an indication as to whether a document is an original or a counterfeit.

3.7.3 Printing and validation of self validating security documents

Security documents can have multiple fields or areas, each of which contains information that is perceptible in more than one way. One field can contain a visually perceptible image and a digital watermark that can be detected when the image is scanned and processed, another field can contain machine readable OCR text that can be read by both a human and a programmed computer, and still another field can contain watermark data which

can be correlated to the output of a fingerprint reader or apparatus which scans a user's iris. Documents are produced by beginning with a template which defines the placements of elements on the document and the interrelationships between hidden and visual information on the document. The template specifies the placement of elements such as images, photographs, and text and also the interrelationship between information that is visually perceptible to a user of the document and information that is hidden by means of digital watermarks. Different hidden digital watermark data is included in multiple elements of the document. The watermarks in the different graphic elements of the document are correlated to each other and to the visual material on the document. Thus, the document cannot be forged by replacing one element (such as a picture) with a similar element from another document. In order to produce a document defined by a particular template, appropriate pictures, graphics and digital data are extracted from a data bank, and watermark data is embedded in the pictures and graphics as appropriate. The merged digital data is then sent to a printing engine and the final document is produced. An automatic validation system of the present invention reads multiple fields on the document, and it also automatically detects information about the user. The various information is correlated to validate the document.

3.7.4 Method and system for digital image signatures

A signature is inseparably embedded within the visible image, the signature persisting through image transforms that include resizing as well as conversion to print or film and back to digital form. Signature points are selected from among the pixels of an original image. The pixel values of the signature points and surrounding pixels are adjusted by an amount detectable by a digital scanner. The adjusted signature points form a digital

signature which is stored for future identification of subject images derived from the image. In one embodiment, a signature is embedded within an image by locating relative extrema in the continuous space of pixel values and selecting the signature points from among the extrema. Preferably, the signature is redundantly embedded in the image such that any of the redundant representations can be used to identify the signature. Identification of a subject image includes ensuring that the subject image is normalized with respect to the original image or the signed image. Preferably, the normalized subject image is compared with the stored digital signature.

3.8 Digimarc Products

3.8.1 Digimarc ImageBridge

'Digimarc ImageBridge' digital watermarking is the industry leading digital watermarking solution that allows digital image owners to place imperceptible codes into their digital images and photographs. The software is designed to work with Adobe Photoshop 7. The technique is integrated into image editing applications for simple digital watermark embedding and reading. When an image is opened in a Digimarc-enabled image editing application, the image is automatically checked for a Digimarc watermark. If it is found, then a copyright symbol, ©, is added to the image title bar to show that the image is being protected. Thus the image reader allows to

- Determine the copyright owner of an image
- Contact the owner directly for image licensing
- Enable licensing and e-commerce opportunities

The Digimarc ImageBridge^{TM} Digital Watermarking Software Development Kit (SDK) available through a flexible C/C++ callable application programming interface (API). The Digimarc ImageBridge Digital Watermarking SDK provides three basic function calls: Detect Watermark, Read Watermark and Embed Watermark. Detect Watermark offers a quick means of determining if a digital watermark is present in an image, typically in less than one second regardless of image size. Read Watermark retrieves the contents of a Digimarc digital watermark found in an image. Embed Watermark embeds a Digimarc digital watermark in an image. The developer controls the three functions through two basic structures. The first describes the location of image data in memory and the second describes the data that is to be embedded in the image. This simple interface provides quick integration and maximum developer control. For maximum flexibility, applications provide pixel data to embed, read or detect functions.

3.8.2 Digimarc MarcSpider

This is another image tracking technology which offers the following advantages:

- It scans the most widely used public areas of the Internet for Digimarc ImageBridge^{TM} digitally watermarked images and reports details on when and where the images are found.

- It also maintains an archive of images found, searchable by date range, so no information is missed out.

Digimarc MarcSpider image tracking uses hundreds of individual spiders, or search engines that look through the web for images which contain Digimarc ImageBridge^{TM} digital watermarks. To effectively cover the Inter-

net, MarcSpider makes some decisions along the way. In general, MarcSpider will visit any given Internet page in its database once a month and check if that particular page has changed. If it has, MarcSpider will visit it more frequently - increasing the frequency of its visits. However, it is unable to access pages that are password-protected.

3.8.3 Digimarc Excalibur copy detection technology

This is part of the security class family of digital watermarking technologies. The copy detection watermark is designed to be used in a variety of printed media where the identification of an original versus a counterfeit is desired. The technology has the following characteristics:

- It detects counterfeit print when reproduced on colour copiers, consumer flat bed scanners or commercial grade drum scanners.
- It can be used to detect re-origination
- It cannot be seen or read without use of proprietary hardware and software
- It carries data for source tracking
- It is secured so that only authorized parties can embed or read the mark. Every customer is assigned a unique digital watermark.

3.8.4 Digimarc Excalibur Secure Authentication Technology

The secure authentication mark has the following characteristics:

- It is designed to survive reproduction and is applicable to printed material without change in material cost or production floor workflow
- It is secured via protocol keys so that only authorized parties can embed or read the mark.
- It carries data for source tracking
- It cannot be seen or read without use of proprietary hardware and software

3.9 Security Features

1-D and 2-D Bar Codes - Bar codes allow data to be stored on cards. 1-D conforms to AAMVA standard. 2-D conforms to AAMVA and PDF47 standards.

Altered (or Modified) Fonts - Slight modifications of text characters are obvious only to trained personnel.

Biometrics - Unique features of the card carrier (fingerprints or iris patterns) are tied to an identifier, often a number, and recorded in a database. The identifier is encrypted and printed on the issued card. The system scans the card carrier's unique identifier and matches it against the information in the database to confirm the card carrier's identity. Biometrics are used for both 1-to-1 authentication at the time of renewal/registration and 1-to-many verification of the new applicant's unique identity as compared to the database of existing document holders, which reduces the opportunity for obtaining valid identity documents under false pretenses. As an integrator of secure identification technologies, Digimarc ID Systems favors no specific biometric

Figure 3.10: Customized Digimarc ID solution example[Used with permission from Digimarc]

technology and recommends solutions based on the client's technology and political environments. Digimarc ID Systems, in conjunction with Visionics, has deployed a range of biometric solutions.

Fine-Line Printing - A pattern of fine lines, similar to those found on currency, can be placed on documents and photo backgrounds. This feature can thwart attempts at photocopying.

Ghost Image - A faint photo image covers printed data, making it virtually impossible to alter and requires no special equipment for verification.

Microprinting - Used on currency, the resolution required to create this

printing on the base material is far beyond that of any known photocopier. Often combined with misspelled words, this feature also prevents misuse of base material and can be read with a small (10) power magnifier.

Optically Variable Ink - Images or text are printed on a card's inner laminate with gold or silver optically variable ink. This printing appears and disappears with the angle of viewing and cannot be photocopied or altered without destroying core laminate. It requires no special equipment for verification.

Signature Area - Signature is captured electronically and printed digitally.

Split Fountain - This refers to the use of a colour degrade that cannot be colour copied.

Unique or Sequential Numbering - Numbers can be added for authentication and to keep track of production materials.

Digital Watermarking - Digital watermarking technology allows issuers to embed digital codes in cards. These embedded codes can be imperceptible to humans, but read by image-capture devices enabled with special reader software. Unlike overt data carrying features, such as barcodes, there is little to alert the forger that a security feature exists. Even if the forger is aware of the feature's existence, it is virtually impossible for the counterfeiter to manipulate that feature in that document or replicate it in another. The covert digital watermark carries a packet of digital data used to authenticate the card and enhance cardholder verification; data, such as a document number, expiration date, birth date, or other data specific to the bearer. The software-based reader can easily be added to primary or secondary inspection. Exam-

ination of documents that use digital watermarks can reveal alteration or forgery by comparing the digital watermark data to other card data.

Overlapping Images - Overlapping images prevent tampering or image substitution.

Multicolour UV - Multicolour images made from red, green and blue ultraviolet ink are visible only when viewed under an ultraviolet light source.

Optical Two-Colour Printing - Two-colour optical printing causes an image or text to shift from one colour to another with the viewing angle. This feature is destroyed with tampering and it cannot be colour copied.

Pattern Printing - This feature incorporates both visible and ultraviolet printing of a pattern or logo on the inner surface of the laminate. Again, it cannot be reproduced with photography or copying.

Ultraviolet Ink - Images printed in UV ink on the laminate are visible only under an ultraviolet light source. Any alteration destroys the feature.

Security Indicia - This feature incorporates words or symbols that are concealed on a document and appear only when a specially grooved plastic viewer is moved across it. This is an inexpensive authentication method which prevents fraudulent use of base material.

Magnetic Stripe - Magnetic stripes can be encoded with information, such as demographic data. It conforms to AAMVA, ANSI and ISO standards.

Writeable Back - A Specified area allows an ID to accept ball-point pen ink.

Microtaggant Particles - Microscopic particles are colour-coded by customer, and can only be seen under a microscope. Optional UV fluorescent feature offers a "quick check".

3.10 Microbar Security: An Overview

There are many ways of making it difficult for counterfeiters to produce counterfeit credit cards or high value paper such as bank notes, vouchers, travellers cheques, tickets or bonds. However, many times people handling the cards or the bank notes do not inspect them carefully and do not have the expertise or equipment necessary to detect forgeries.

'Microbar' is an innovative way of using conventional printing techniques to give secure machine readable validation. It utilises fractal geometry to encode data into the background of the document. Once the data to be included in the Microbar has been decided, a mathematical formula turns the data into a seemingly abstract pattern or dots. A second formula and key is then used to blend this pattern into the background image of the document being marked. The document can then be scanned using a conventional scanner and the Microbar can be read and decoded. If it is forgery, the scanner, using the key hidden in the pattern telling it how to unlock the data, cannot decipher the Microbar and the user is alerted of the problem.

The main benefits of the Microbar are as follows:

- Microbar is invisible to the naked eye and can be easily hidden in the background of an image, such as the watermark on bank notes and certificates.
- Because it uses standard printing and reading technology, Microbar can effortlessly be put on anything from bank notes and plastic cards to

shipping containers.

- Authentication is via a simple, low cost scanning technology which is able to differentiate between the presence of an encrypted signature on an original from a very good copy with extremely high levels of confidence.
- Microbar is far more secure than any other encryption method known, and from a mathematical point of view, it is effectively impossible to crack the coding system used.

3.10.1 What is Fractal Geometry?

Martin J. Turner et al[63] in their book on 'Fractal Geometry in Digital Imaging' gives a very good account of the mathematical analysis of Fractals. Fractal geometry is being widely used in diverse fields of physics and mathematics. This technique is particularly suited to model natural objects. Fractal geometry dates back to the late 1800s and early 1900s. Mathematical curves, such as the Peano Curve can be called as a mathematical fractal.

The term 'fractal' was coined by Mandelbrot in 1975 to describe the irregular structure of many natural objects and phenomena. A fractal description may be a deterministic fractal, for example the Von Koch snowflake[63]. In real life the description is statistical when discussing random fractals that are more useful in modelling textured images.

Fractal Mathematics:

The Fourier Fractal Dimension (D_F) of an image is given by

$$D_F = \frac{8 - \beta}{2}$$

where β is the spectral exponent.

A synthetic fractal is created by filtering white noise of the required size with a low pass filter, q. If Q is the Fourier Transform of q, then for two-dimensional processing

$$Q(k_x, k_y) = |k|^{\beta/2}$$

where $|k| = \sqrt{k_x^2 + k_y^2}$ and $\beta = 8 - 2D$. A synthetic fractal landscape is created on the principle of forming two-dimensional fractal noise. The same principle can be applied to creating synthetic fractal signals. The four stage process of creating fractal noise is described below:

1. Create a random Gaussian array with zero mean and unit variance. Create a uniform random sequence (normalised) of the same size.

2. Calculate real and imaginary parts of the noise using $N_i = G_i \cos 2\pi U_i$ and $M_i = G_i \sin 2\pi U_i$, defining G_i as amplitudes and U_i as phases.

3. Filter N_i and M_i with $W_i = \frac{1}{k_i^{\beta/2}}$ to create N' and M', where $\beta = 8-2D$ and $|k_i| = \sqrt{k_x^2 + k_y^2}$.

4. Apply Inverse Fast Fourier Transform (IFFT) to obtain $n_i = Re\left(\hat{\mathcal{F}}^{-1}\{N' + iM'\}\right)$ where $\hat{\mathcal{F}}^{-1}$ is the inverse Fourier transform.

The theory of fractal geometry and its algorithms are used extensively in the Microbar two-dimensional watermarking schemes. These schemes are described in depth in Chapter Five. The detailed mathematical analysis of synthetic fractals can be obtained from [63].

3.10.2 Microbar: The Concept

There are numerous methods available to protect printed artefacts from counterfeiting. Many of these however rely either on the incorporation of an

overt feature (which might otherwise not be desired for space/design reasons) or require the use of special inks, substrates or taggants linked with bespoke reader technology and/or expert intervention to validate originals.

The *Microbar*^{TM} technique integrates seamlessly into existing printing infrastructures to provide a secure anti-counterfeit mark which can be totally camouflaged yet will allow fast automatic verification without recourse to expert intervention.

3.10.3 Microbar: The Product

Microbar^{TM} is an internationally patented invention which makes use of a multi-dimensional statistical encryption to give powerful, non-repeating patterning that is in itself a digital code. This technique may be applied either to the whole artefact or specific areas where the marking is disguised as an incidental or artistic feature. This option ensures an unobtrusive coding and yet provides for integrity of the coding information so that it may still be validated even if damage or defacement of the surface has occurred. *Microbar*^{TM} can be used in a variety of printing techniques in either black & white or colour.

Easy to Use - *Microbar*^{TM} is a validation system that uses standard printing and reading technology. Statistically robust results can be achieved using scanning resolutions as low as 100 dpi.

Secure - *Microbar*^{TM} protection cannot be reversed. However detailed and apparently 'perfect' a copy is the validation system can easily distinguish it from the original. Also it is impossible to work backwards from an original printed document to recreate the original *Microbar*^{TM} file in order to counterfeit the document by reprinting.

Covert - A graphics file to be protected is modified by *Microbar*^{TM} but appears visibly unchanged when printed.

Low Cost - Once the graphics file has been encrypted normal printing methods are used. No special inks, substrates or taggants are required.

Unique - This level of security has never been possible before using standard printing and scanning.

3.10.4 Physical Testing and Security

Such non-destructive validation is very attractive for high value bonds and documentation. *Microbar*^{TM} is being developed in partnership with Debden Security Printing, a wholly owned subsidiary of The Bank of England. Physical security of the unique Microbar system can be ensured at the Bank's printing works where all the Bank of England's banknotes are printed.

3.10.5 Other Uses

In addition to simple authentication, any suitable scanner is capable of reading and extracting digital information from the encrypted signature on the surface of the marked card, label or document. This covert information can be processed either by itself in the case of documents/vouchers/labels or used in conjunction with magnetic stripe/microchip data in the case of credit cards to validate the medium. *Microbar*^{TM} has the capability to hide and contain a significant amount of data and so has the potential to act as a covert tracking mechanism for brand owners concerned as much with diversion as with counterfeiting.

3.11 Conclusion

Various Watermarking algorithms in use today have been reviewed and briefly referenced in many instances. It has been found that the watermark schemes in the spatial domain, LSB etc. are much faster than those in the transform domain like DCT, etc. However, in the spatial domain they are less robust to compression and geometric distortions. Watermarks using LSB can easily be removed using compression and low pass filtering.

Digital watermarking is still a very active research area and by far a mature field. Currently, there is research in three directions. Watermarking algorithms that are more content dependent are being investigated to combat copy attacks. The copy attack copies a watermark from one image to the other without knowledge of embedded systems or cryptographic keys used. A watermark that is dependent on the contents of the data being watermarked will resist this kind of attack.

The second direction in which digital watermarking research is progressing is the digital video area. Many research activities are now directed to low bit-rate video watermarking. Also some research activities are now being conducted for watermarking systems that watermark both video and audio parts of an audio visual data to protect against alteration of one of the components without interfering with the others. The final direction of watermark research deals with providing potential users of watermark systems information about the system's performance. Research is being carried out to develop an internationally recognised watermark benchmarking system. However, there is no significant discussion about whether the watermarking system should be standardised.

Chapter 4

The Lippmann Optical Variable Device

4.1 The Theory of the Photographic Process

4.1.1 Introduction

This chapter is concerned with the basics of the photographic process, the exposure of the object to the photographic material, formation of the latent image and the development to give a stable permanent image. It also explains the Lippmann photographic technique and its practical applications. A mathematical model of the emulsion has been developed and extensive laboratory tests have been conducted to prove the principle and its suitability for security applications.

Photography is basically a two stage process involving firstly the formation of an image by physical means and secondly the preparation of a permanent copy of this image by chemical means. The image may be formed by the lens of a camera, by shadows cast by X-rays or gamma rays, or by

the trace of a moving electron beam, or by other invisible electromagnetic radiation such as infrared or ultraviolet radiation.

The photographic material used is the silver halide material similar to the one used in holography. The main advantages of this material are that it has high sensitivity in comparison with other photographic materials available and it can be coated on both film and glass. It also has some drawbacks such as it is absorptive and requires wet processing.

A silver-halide recording photographic material is based on one type or a combination of silver-halide crystals embedded in a gelatin layer, commonly known as the photographic emulsion. This emulsion is coated on a flexible or stable substrate material such as film or glass.

The basic sequence in obtaining images by means of silver halides is as follows:

- Exposure to electromagnetic radiation to form an invisible latent image - the exposure H is defined as the incident intensity E times the time t of exposure of the recording material. If the intensity is constant during the whole exposure time, which is usually the case, then we have the relation: $H = Et$. In reality, the exposure necessary for obtaining a certain density in the developed material is not constant but depends on time t. For very short exposures at high intensities E, as well as for very long exposures at low intensities, H has to be strongly increased to get the same density as the one required for the optimal values of E and t. As the intensity increases, more absorbed photons are required per grain to produce the same density in the developed material as when compared with exposures at lower intensity levels.

- Amplification of the latent image by development to render it visible and subsequent processes to make the image stable and permanent:

The exposure of a silver halide material, will after processing - i.e. development and fixing result in a certain optical density of the material, as silver is created during the development. The higher the exposure of the material, the higher the density will be upto a certain limit. Optical density is a useful characteristic of the light-absorptive property of a particular region of a photographic negative. The greater the density, the less light is transmitted.

Due to the fact that the developed silver halide emulsion consists of millions of discrete silver particles, it is obvious that when a normal photographic image is enlarged enough, the areas which appeared homogenous to the unaided eye will now show a granular pattern. The smaller the grain sizes, the higher is the contrast in the image. Silver halide crystals of the conventional photographic film can be larger than one micrometer, which results in high granularity.

The resolving power of the photographic material is another important property, which gives the measure of the ability of the material to record fine detail. It is the ability of a photographic material to maintain in its developed image the separate identity of parallel bars when their relative displacement is small. Normally the resolving power of photographic materials is tested by using a resolution test chart. The highest number of lines per millimetre that can be resolved in the emulsion corresponds to the resolving power of the tested material.

4.2 Lippmann Photography and Its Principles

4.2.1 Introduction

In 1886, Gabriel Lippmann (1845-1921), professor of mathematical physics at the Sorbonne in France developed a general theory of recording colours as standing waves in a light sensitive emulsion. In 1891, he developed the Interferential photography, also called interference colour photography, which was the first direct technique to record colour photographs[87]. He also developed the first theory of recording monochromatic and polychromatic spectra in a black and white emulsion using the principles of Fourier mathematics. The new technique however, was not very effective for colour photography since it was complicated and the exposure times were too long for practical use. The difficulty in viewing the photographs was another contributing factor, in addition to the copying problem, which prevented Lippmann photography from becoming a practical photographic colour-recording method.

However, a new optical security application of this technique has been investigated. Currently, this type of technique can be applied as a unique security device on security documents, like passports, identification cards, credit cards, driving licences etc. This method offers a very high degree of security and has many advantages.

To list a few of these advantages:

- The Lippmann OVD has a very high archival stability.
- The Lippmann OVD is Bragg sensitive, which means it changes its colour depending on the angle of illumination and observation[87].

- The Lippmann OVD cannot be copied by conventional colour photography nor can it be copied on colour copy machines.

High-resolution panchromatic recording materials suitable for Lippmann photography are currently available. In particular, the use of colour photopolymers from DuPont has been investigated for modern Lippmann Photography. Since the colour photographs contain no dyes or pigments their archival stability is very high. The photopolymer material requires dry processing and no expensive equipment, such as lasers, are needed to explore this photographic recording technique, only a modified camera (described later) has been used.

4.2.2 Principle of Lippmann Photography

The recording material has to be of a low light sensitivity due to the demand for high resolving power for recording Lippmann photographs. The photosensitive emulsion coated on Lippmann plates is brought in contact with a highly reflecting surface[89]. Lippmann used mercury in contact with the emulsion. This mirror reflects the light into the emulsion, which then interferes with the light coming from the other side of the emulsion. Stationary standing waves of the interfering light produce a very fine fringe pattern throughout the emulsion with a periodic spacing of $\lambda/(2n)$ (λ is the wavelength of light in air and n is the refractive index of the emulsion). The colour information concerning the object is recorded in this way. For example a large separation between the fringes in the emulsion indicates that the recorded wavelength is located at the red end of the spectrum, however more closely spaced fringes indicate a shorter wavelength, such as green or blue. This is only applicable when monochromatic colours are recorded. When the developed photograph is viewed in white light, different parts of the recorded

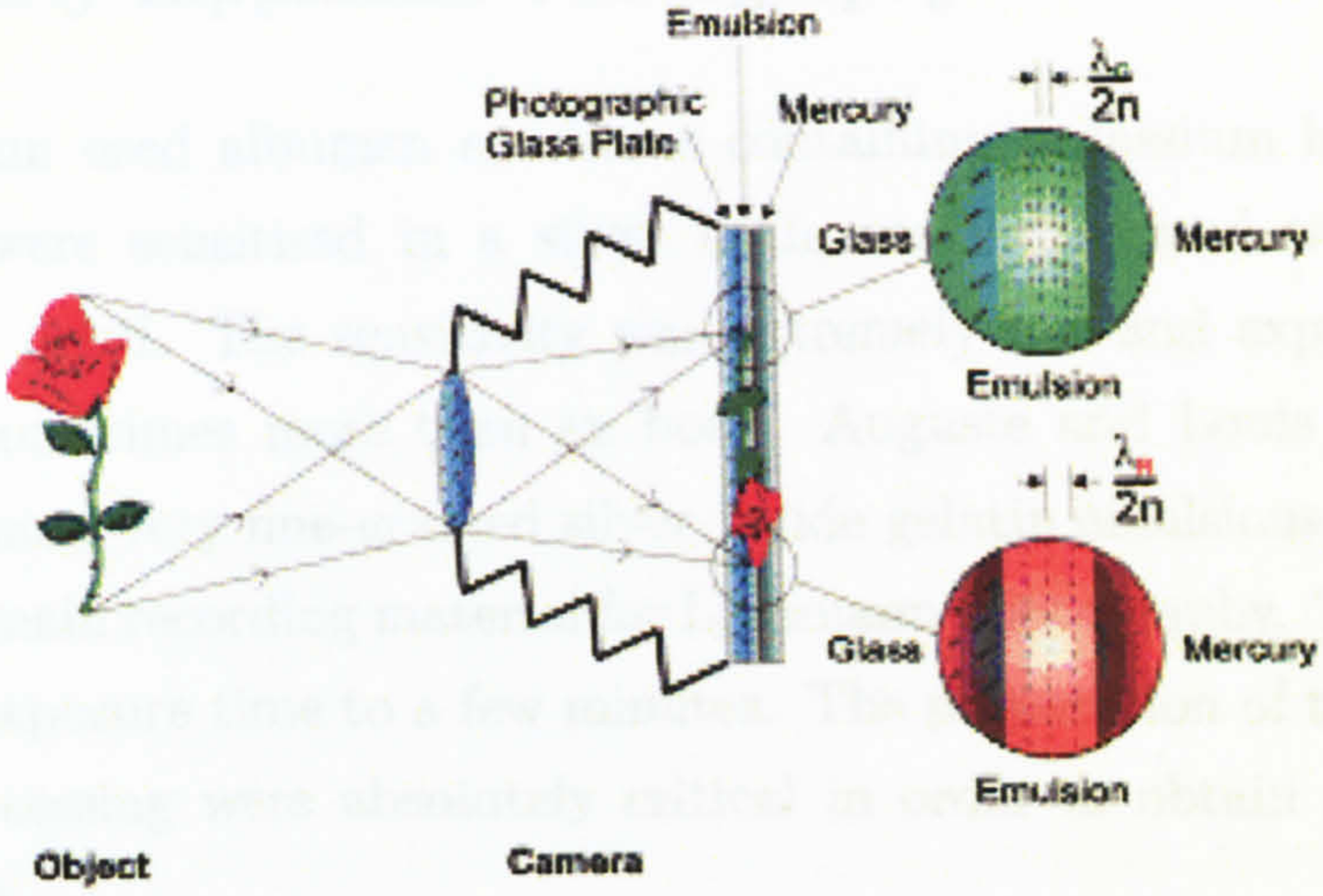


Figure 4.1: Adapted from [93] Principle of Lippmann Photography

image produce different colours due to the separation of the recorded fringes in the emulsion. The light is reflected from the fringes, creating different colours corresponding to the original object. There is a severe requirement on the resolving power to record the fringes separated in the order of half the wavelength of the light. The processing of these plates is critically important, changing the separation between the fringes would create wrong colours. To observe the correct colours, the illumination and the observation must be at normal incidence. The colour of the image changes with the angle of viewing called as iridescence, an important attribute of the Lippmann photograph as a document security device.

Lippmann photography involves no phase recording; the recorded interference structure is a result of phase-locking the light by the reflecting mirror.

4.2.3 Early Lippmann Photography

Lippmann used albumen emulsions containing potassium bromide[89]. The plates were sensitised in a silver bath, washed, flowed with cyanine solution and dried. The sensitivity was extremely low and exposure times were high, sometimes more than an hour. Auguste and Louis Lumiere in Lyon introduced very fine-grained silver-halide gelatin emulsions which later became the main recording material for Lippmann photography. These plates reduce the exposure time to a few minutes. The preparation of the emulsion and the processing were absolutely critical in order to obtain good colour photographs.

Most experimenters used developers based on pyrogallol and ammonia, which were formulated to suit the particular emulsion. A surface developer performed well since no image information is located deep inside the emulsion. Development time was one to three minutes. Sometimes the image was fixed. The Lumiere brothers recommended immersion in a potassium cyanide bath[89]. However, sodium thiosulfate bath was also used. Often it was recommended not to fix the developed image, since that would change the thickness of the emulsion and, thus, change the colour of the image. Rarely was the plate intensified, redeveloped or even bleached. Ives tested other developers based on amidol and hydroquinone and used bleaching to create phase gratings which came out more brilliant and narrow band than pyrogallol-developed ones[89].

4.2.4 Modern Lippmann Photography

Single-beam Denisyuk reflection holography shows similarities to Lippmann photography[89]. In both the cases an interference pattern is recorded in a high-resolution emulsion and the principle of Bragg diffraction applies

to both. The fundamental difference however, is that in the Lippmann case, there is no phase recording involved; the recorded interference structure is a result of phase-locking the light by the reflecting mirror. In holography, the phase information is actually recorded, being encoded as an interference pattern created between the light reflected from the object and a coherent reference beam. A Lippmann photograph therefore can be regarded as a reflection image-plane hologram recorded with light of very short temporal coherence. The reference wave is a diffuse complex wave front, the mirror image of the exit pupil of the recording lens.

The recording of monochromatic light in a Lippmann emulsion is very similar to recording a reflection volume hologram. A broadband polychromatic spectrum, such as a landscape image, is however different. In this case, the recorded interference structure in the emulsion is located very close to the surface of the emulsion in contact with the reflecting mirror. Thus, an emulsion thickness of only a few micrometers is needed and actually preferred.

4.3 Recording Materials for Lippmann Photography

4.3.1 Introduction

New and improved recording materials combined with special recording and processing techniques have made it possible to develop the new OVD method. The optimal recording material for Lippmann photography has to be isochromatic to give a correct colour recording. Most holographic materials are not perfectly isochromatic and therefore colour correction filters

are sometimes used for recording Lippmann photographs to obtain the right colour balance.

The two holographic recording materials suitable for Lippmann photography are panchromatic ultra-high-resolution silver halide materials and panchromatic photopolymer materials. Silver-halide materials require wet processing. Here the ultra-fine-grain panchromatic emulsion (PFG-03C emulsion) from Slavich coated on film has been used. Some characteristics of the Slavich PFG-03C material are presented in Table 3.1.

Silver halide material	PFG-03C
Emulsion thickness	7 nm
Grain size	10 - 20 nm
Resolution	$\sim 10000 lp/mm$
Blue sensitivity	$\sim 1.0 - 1.5 10^{-3} J/cm^2$
Green sensitivity	$\sim 1.2 - 1.6 10^{-3} J/cm^2$
Red sensitivity	$\sim 0.8 - 1.2 10^{-3} J/cm^2$
Color sensitivity peaked at:	633 nm, 530 nm, 450 nm

Table 4.1: Characteristics of the Slavich PFG-03C emulsion

The panchromatic holographic photopolymer materials from E. I. du Pont de Nemours & Co. require only dry processing (UV-curing and baking) which makes them particularly suitable for Lippmann OVD security products. Although, less sensitive than the Slavich silver-halide emulsion (which is also slow, by modern photographic standards), it has its special advantages of easy handling and dry processing.

A Lippmann photograph is recorded in photopolymer materials in the following way. The photosensitive polymer layer has to be thin, of the order of a few micrometers only. It is coated on a flexible transparent base and a

special type of reflecting foil is laminated on top of the photosensitive polymer layer under safe light, making sure it is in absolute perfect contact with it. In the experiments, the panchromatic photopolymer material (HRF-700X071-3 film) having an emulsion thickness of about 2 to 4 μm is used. As a reflecting surface, silver sputtered (800Å) polyester film without the standard anti-dust oxide (InO) top layer is used. Sometimes a colour correction filter may be needed in front of the camera lens to obtain correct colour balance.

4.3.2 Practical work on Slavich silver-halide emulsions using various test targets

These tests proved that the Lippmann technique works very well. It involved exposing the test targets for different times in order to get the right colour balance and developing them using the Lumiere developer. Tests were also carried out using the Birenheide HRT (German) and Slavich plates. It was found that the Slavich plates gave better results. These tests were carried out indoors under artificial lighting conditions. An example of the test target used is a sample passport shown in Figure 4.9.

Tests have also been conducted using the GP8 developer and a comparison from the results shows that the Lumiere developer might be better under indoor conditions. A few tests were also conducted under outdoor conditions with varying degrees of sunlight.

Various other test targets were used, for example the Macbeth Colour Checker (Figure 4.2) which is a very good example of the sharpness and quality of the colours that can be achieved with this technique.

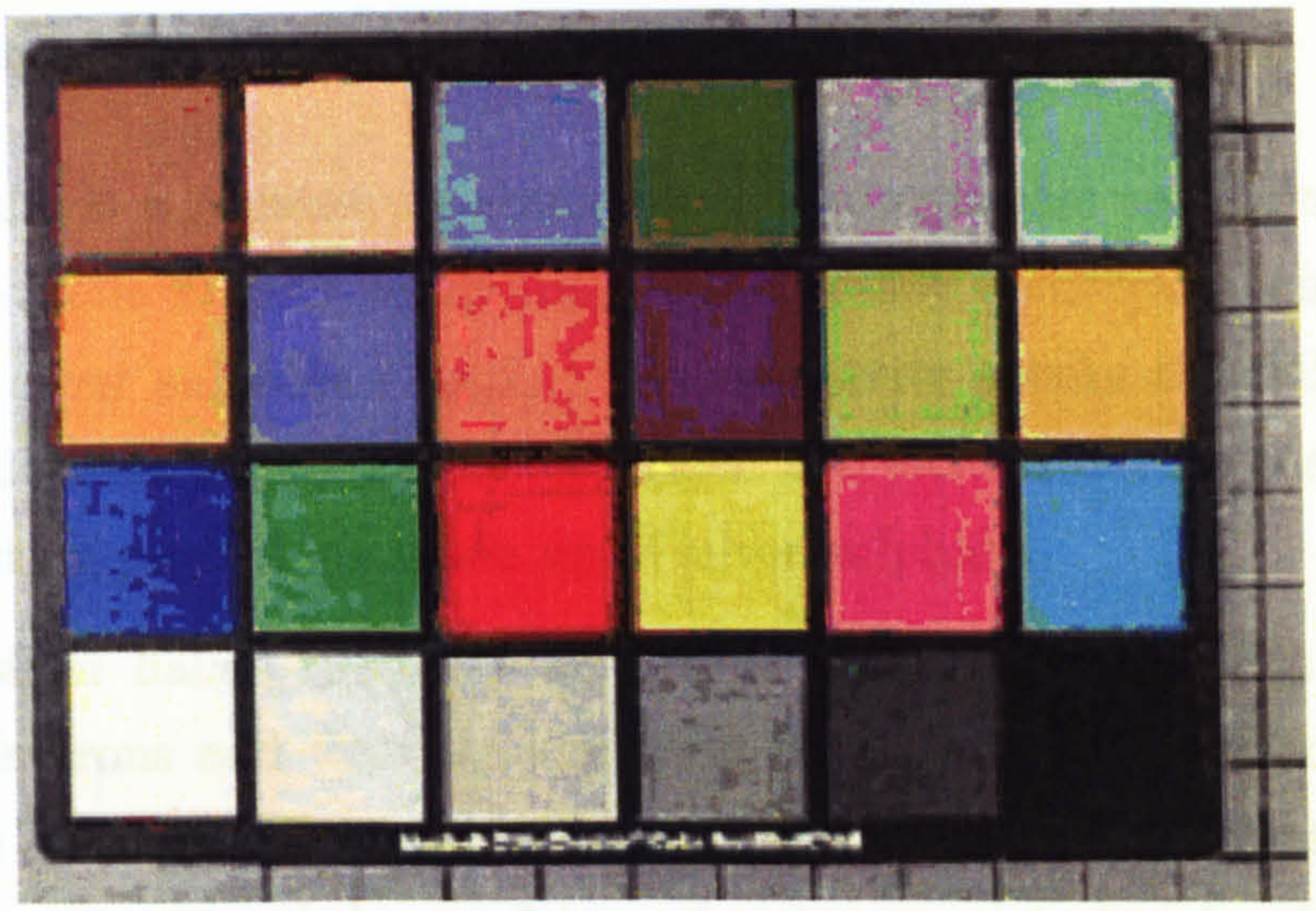


Figure 4.2: Macbeth Colour Checker

The Lumiere developer consists of the following:

Solution A		*Solution B*	
Pyrogallol	1g	Potassium bromide	10g
Water (distilled)	100ml	Water (distilled)	100ml

When viewing Lippmann photographs, the reflection from the gelatine surface of the emulsion can cause colour distortion as a result of the phase shift in the gelatine/mercury interface during recording[94]. In addition, the specular surface reflection has to be separated from the image in order to see it clearly. This can be done by attaching a wedged (angle about 100°) glass plate on top of the emulsion using index-matching glue, e.g. Canada balsam. The back of the photograph is painted black, covered with black paper and the edges sealed. The photographs could also be viewed enlarged, by projecting the reflected image using a special projector.

Formation of the latent image and Development[96]

Gelatine is a necessary component of photographic emulsions because it contains labile sulphur compounds, which easily decompose when heated producing silver sulphide (Ag_2S). The sensitivity specks that silver-halide grains exhibit on their surface and which are formed during the emulsion manufacturing process are made up of silver sulphide.

The silver halide crystal is an n type photoconductor with a valence band of electrons and with a conduction band in which injected electrons are free to migrate throughout the crystal until trapped by a lattice defect. During the exposure of the emulsion, photons are absorbed by the crystals. When a photon of sufficient energy is absorbed, an electron from the crystal is promoted to the conduction band, leaving behind a positive hole which is a free halogen atom:

$$Ag^+X^- \text{ (silver halide crystal)} + h\nu \rightarrow Ag^+X^0 + e^-$$

where ν is the frequency of light (photon energy). The photogenerated hole is trapped at a surface sensitivity site by partial S^{2-} charges from the adsorbed silver sulphide specks. This results in a positively charged Ag_2S^+ particle, which dissociates into AgS and Ag^+.

$$Ag_2S^+ \rightarrow AgS + Ag^+$$

The silver ion then attracts the photogenerated electron to form a silver atom.

$$Ag^+ + e^- \leftrightarrow Ag^0$$

Typically, the isolated silver atom has a life time of about one second. In order to create a sublatent image speck on the silver halide crystal, where a

diatomic silver molecule is formed by the process of nucleation, a second silver atom is needed to combine with the first silver atom during its life time. The sublatent image speck grows larger with further photon absorption, resulting in photogenerated electrons. The latent image is usually regarded as a collection of a few silver atoms at one site produced by the reduction of silver ions. For all the exposed grains, chemical development will then reduce the entire silver-halide grains to metallic silver. Development however depends on the sensitivity and the size of the grains in the emulsion. A highly sensitised grain requires fewer photons than a less sensitised grain and larger the grain size, higher is the sensitivity. At high intensities, electrons are produced at a rate that there is not enough time for the mobile silver ions to neutralize the trapped electrons. Recombination of electrons and holes may occur and as a result, the latent image becomes inefficient. At low intensities however, if the isolated silver atom has not combined with another silver atom in its lifetime, then it will decompose into an electron and silver ion again. This shows that a long exposure at low light levels becomes an inefficient process for the latent image formation.

Lumiere developer used becomes active only when it is alkaline with activity increasing with the alkalinity. During the development some of the unexposed silver halide grains also get developed and this causes fog in the photographic emulsion. In order to prevent this a 'restrainer' such as potassium bromide is added to the developer.

Practical Work

The work described in this section has been performed solely by the author at the Modern Optics Laboratory, De Montfort University, Leicester. In order to record Lippmann photographs the following camera was used:

Eastman Kodak Co. (Folmer & Schwing Div.) Auto Graflex 4 x 5 inch camera equipped with a Kodak Aero Ektar F.2.5, 178mm lens. The camera can accommodate both film and glass plates. The panchromatic PFG-03c emulsion, coated on 1.5mm thick glass (plate size 4 x 5 inches), was used to record the Lippmann photographs. Here the gelatine-air interface acts as the reflector of light. The plate was inserted in a conventional dark slide with the emulsion side facing away from the camera lens. The plate holder was then inserted in the camera, mounted on a tripod. When the plate is exposed without mercury, the exposure time is slightly increased compared with recording with a mercury reflector.

Similar experiments were conducted using the Slavich film and using a Canon 35mm camera which has been specially modified for Lippmann photography.

To explain the reason why it is possible to obtain a Lippmann photograph without mercury, the difference between a reflection at the mercury surface and a gelatine-air interface has been studied[93]. A node is located at the mercury reflector (an optically thicker medium than gelatine), which means that at the gelatine surface there is a phase shift of $\pm\pi$. However, a crest is located at the surface when the reflection is obtained from the gelatine-air interface, (an optically thinner medium than mercury) which means that as no phase shift occurs in this case, a silver layer will be created at the emulsion surface after development. In the mercury case, the first silver layer is located at a distance of $\lambda/4$ inside the gelatine emulsion. With regard to the second silver layer, it will be $\lambda/4$ closer to the gelatine surface compared with the mercury reflector case. Since the coherence length of ordinary light is extremely short, the difference in distance from the gelatine surface is extremely important. The slightly increased modulation (caused by a higher degree of coherence) in the gelatine-air reflector case can some-

what compensate for the weaker reflection obtained in this case. However, the exposure must be slightly increased to bring the recording up on the linear part of the Hurter-Driffield curve[108]. The problem concerned with the surface reflection being out of phase with the image when viewing a Lippmann photograph is only valid in the mercury case. When using the air reflector, the surface reflection is in phase with the image. However, in this case as well, it can be recommended to add a glass wedge in order to improve the image contrast.

The Slavich emulsion is rather soft, and it is important to harden the emulsion before the development takes place. The emulsion thickness is about 7μm, which is thicker than necessary for Lippmann photography. The interference pattern is recorded in a very thin volume at the top of the emulsion. This area has to be maintained intact after processing, justifying the prehardening step. Emulsion shrinkage and other emulsion distortions caused by the developer must be avoided.

The following bath was used for this first processing step:

Distilled water	750ml
Formaldehyde 37% (Formalin)	10ml (10.2g)
Potassium bromide	2g
Sodium carbonate (anhydrous)	5g
Add distilled water to make	1litre
Processing time	6minutes

Among the old developers, the Lumiere pyrogallol developer mentioned earlier, gave the best results. The development time is about 90 seconds at approximately $18°C$. After a 10-minute wash, the plates are soaked in distilled water that contains a wetting agent. Finally, the photographs are slowly dried at room temperature. The Lippmann photographs contain no dyes or pigments and their archival stability is very high. Broadband or

narrow band Lippmann filters are another possible scientific application of this photographic recording technique.

4.3.3 Photopolymer materials

A photopolymer recording material consists of three parts: A photopolymerisable monomer, an initiator system (initiates polymerisation upon exposure to light) and a polymer (binder). To explain briefly, how it works; first, an exposure is made to the information carrying interference pattern which polymerises a part of the monomer. Monomer concentration gradients, formed by variation in the amount of polymerisation due to the variation in exposures, gives rise to diffusion of monomer molecules from the regions of high concentration to the regions of lower concentrations. The material is then exposed to regular light of uniform intensity until the remaining monomer is polymerised[94].

A typical filter, which has been used, for the present studies has been shown in Figure 4.3.

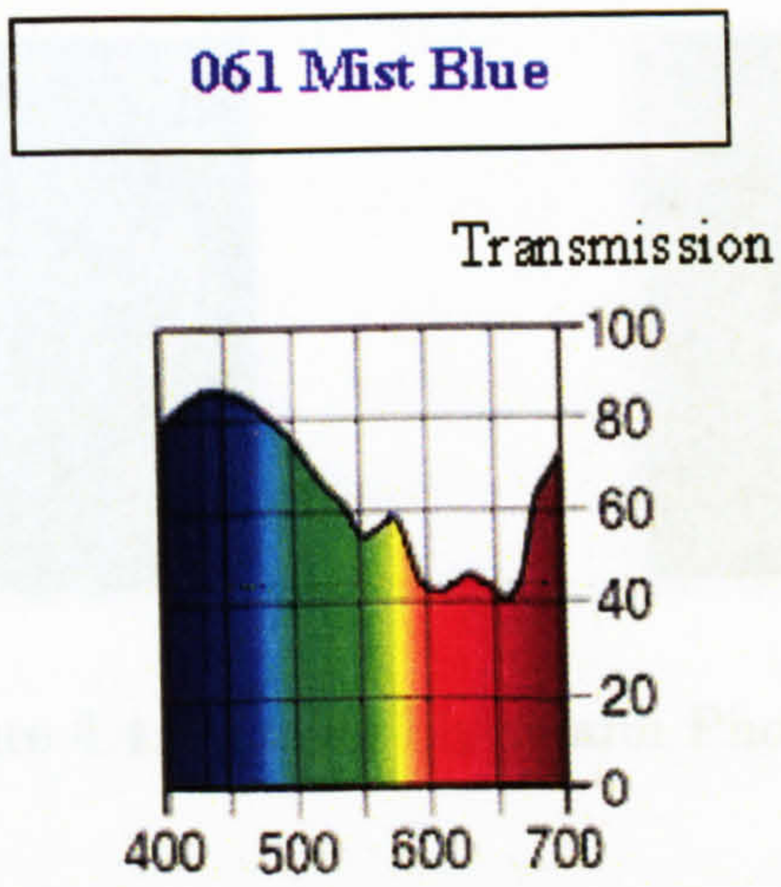


Figure 4.3: Typical filter used in experiments

The above filter was chosen to recreate the colours in the original image after development. However, a catalogue of colour filters are available for colour variations. DuPont recommends about $100mJ/cm^2$ exposure at 350-380nm. After that, the photograph is put in an oven at a temperature of $1200°C$ for 2 hours in order to increase the brightness of the image.

Sample Lippmann photographs are shown in Figure 4.4 and the spectrum recorded in Figure 4.5.

Figure 4.4: Sample Lippmann Photographs

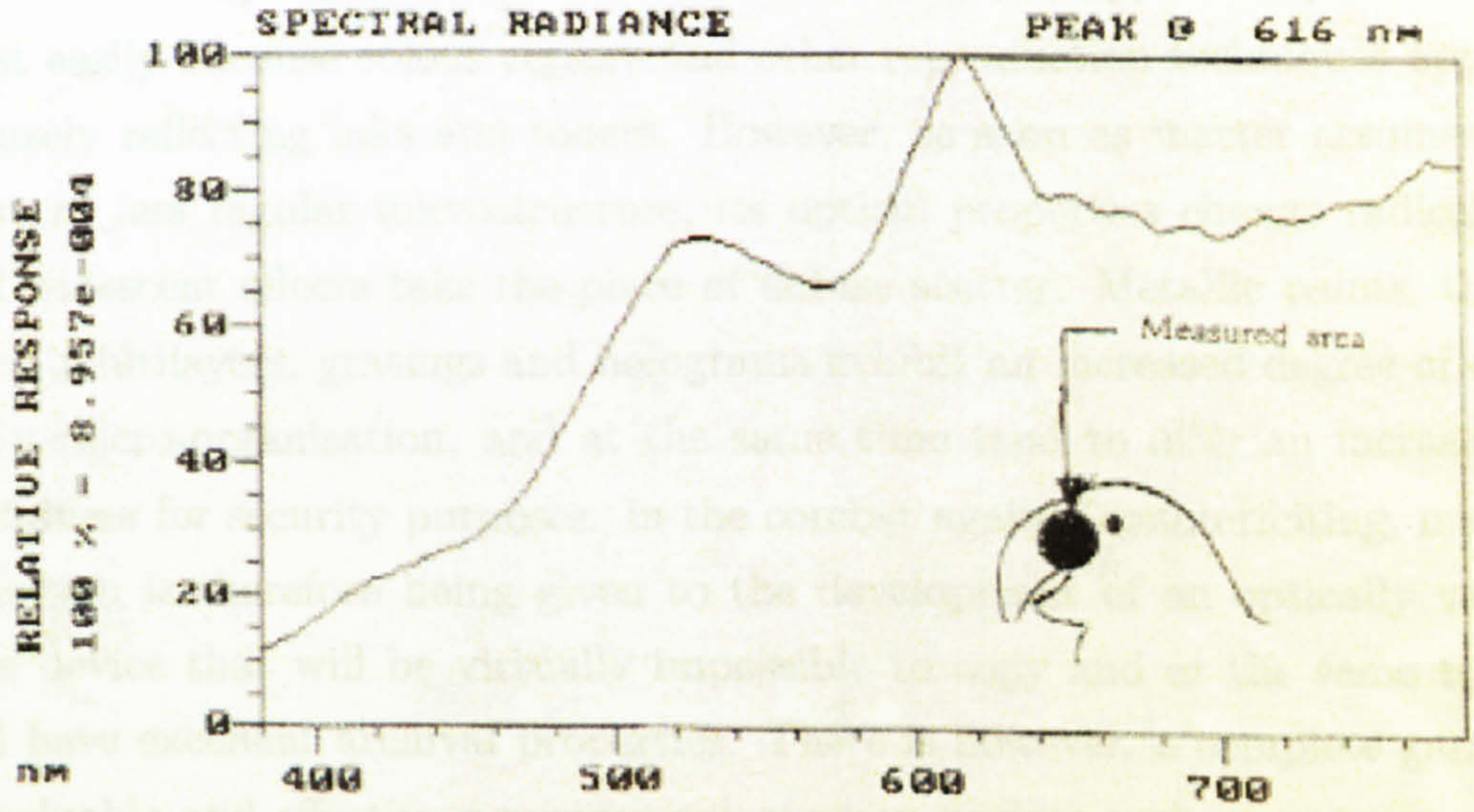


Figure 4.5: Spectrum recorded from Lippmann Photograph of a parrot[87]

4.4 Applications of Lippmann Photography in Document Security

An important key to the subject of document security appears to be the microstructure of matter, which varies between random and highly organised patterns. It appears that the degree of regular organisation of matter is a rough parameter of its value for security features. A highly random structure of matter on a microscopic scale brings about diffuse scatter and renders matter optically invariable; that is; its optical properties become highly independent of the angle of observation. This state of matter appears reproducible most easily because colour copiers and other reproduction techniques apply diffusely reflecting inks and toners. However, as soon as matter assumes a more or less regular microstructure, its optical properties change radically and iridescent effects take the place of diffuse scatter. Metallic paints, thin films, multilayers, gratings and holograms exhibit an increased degree of orderly micro-organisation, and at the same time tend to offer an increased usefulness for security purposes. In the combat against counterfeiting, much attention is therefore being given to the development of an optically variable device that will be virtually impossible to copy and at the same time will have excellent archival properties. There is however, a complete gamut of valuable and effective noniridescent security devices such as watermarks, intaglio printing, metamerism, transitory and tilt images, fluorescence, thermochronism and so on that are already on the market.

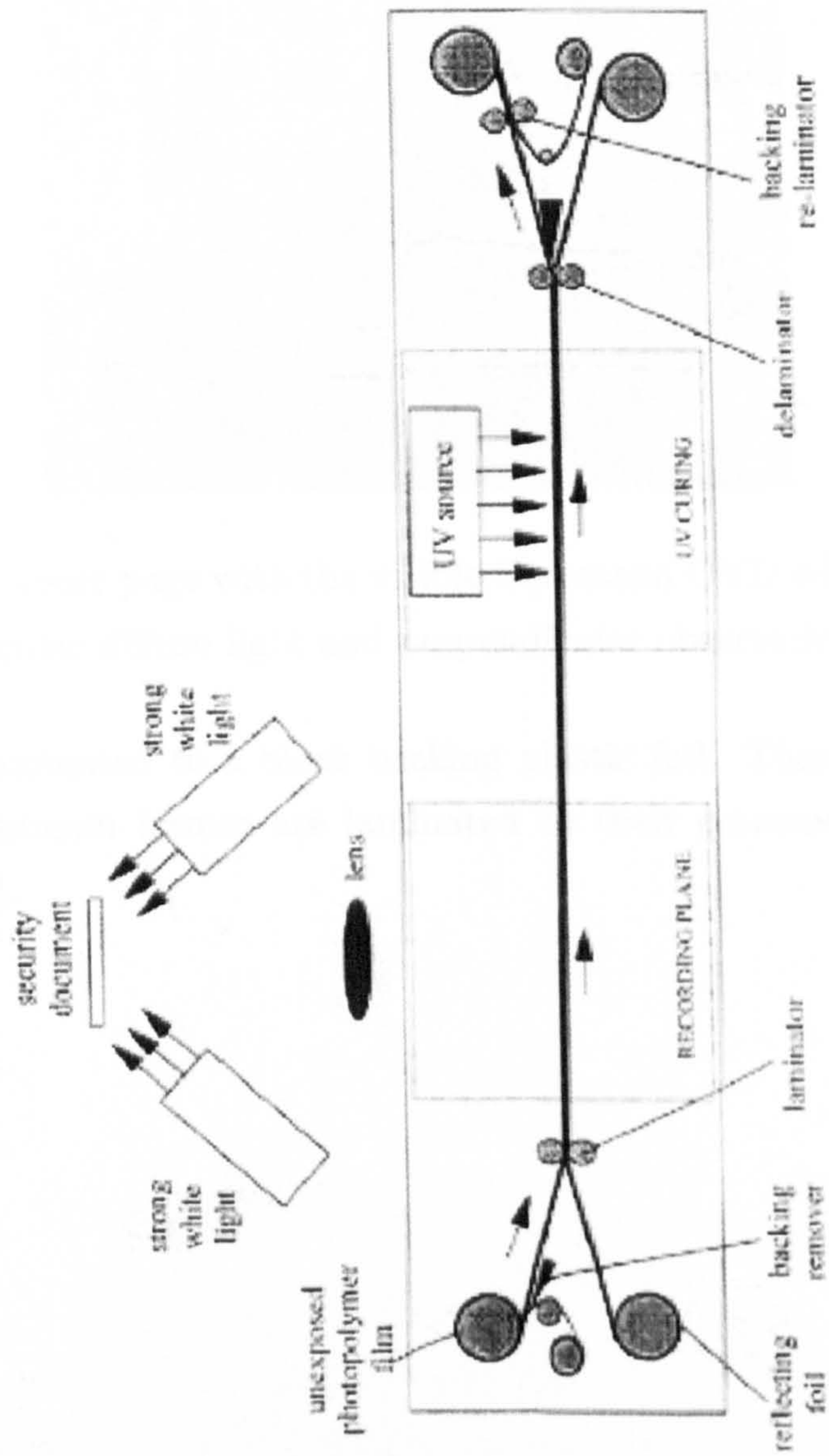


Figure 4.6: Schematic showing method of recording Lippmann security Photographs

Both the recording and the processing can be performed on the same piece of equipment. After being processed and if desired, the recorded im-

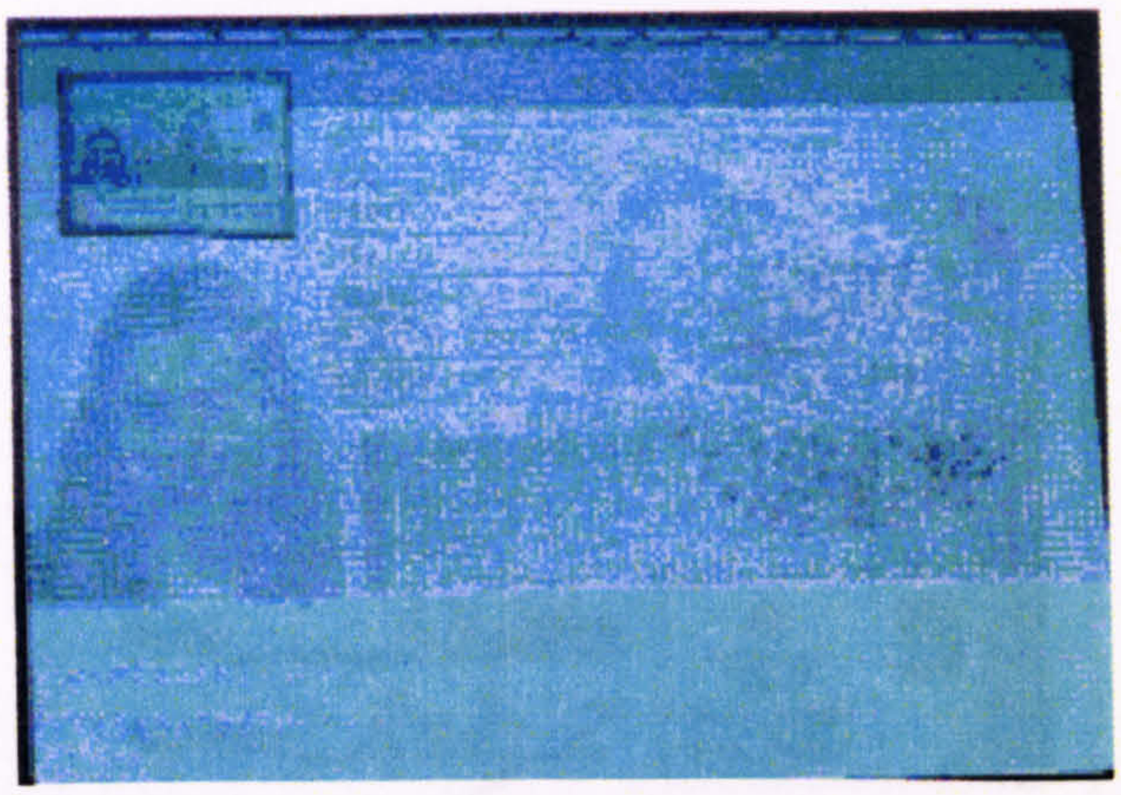


Figure 4.7: Passport page with the visible Lippmann OVD when illuminated with perpendicular diffuse light and perpendicular observation

ages can be laminated to a black backing plastic foil. Then the backed or unbacked Lippmann frames are laminated to their corresponding security documents[89].

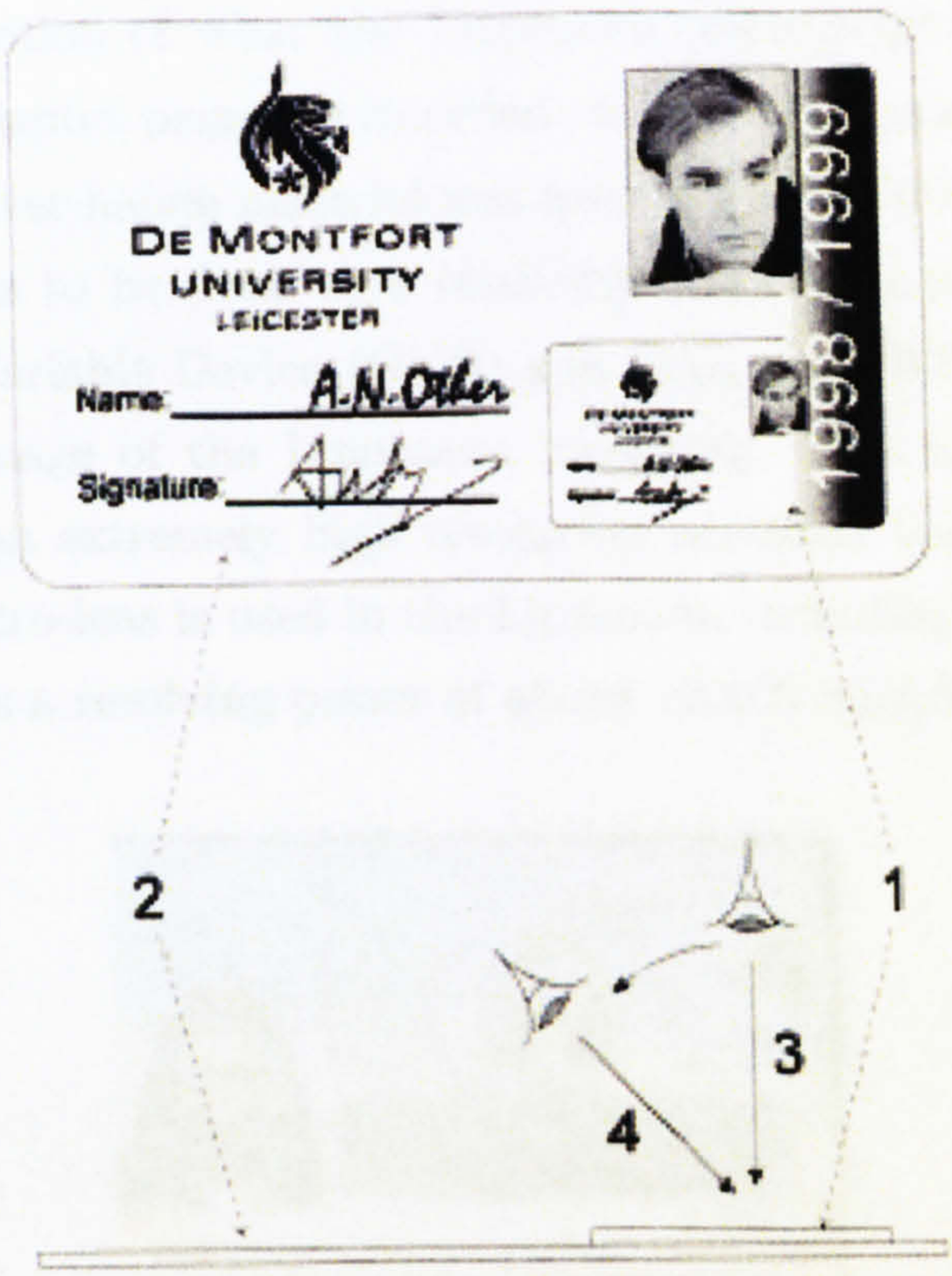


Figure 4.8: Lippmann photograph attached to a security document

In Figure 4.8, a Lippmann photograph (1) is attached to a security document (2), the colour of the image in the Lippmann photograph varies depending on the angle of observation. Perpendicular observation (3) gives the correct colour image; oblique observation (4) shifts the colours toward the shorter wavelengths[89]. Due to the Bragg sensitivity, it is not possible to replace a Lippmann colour photograph with a conventional colour photograph, as mentioned earlier the security document with the Lippmann photograph cannot be copied on a conventional colour copier.

4.4.1 Practical Illustration

As an illustration of what the Lippmann photographic technique can offer, a sample passport page was recorded. In this case, panchromatic ultra-high-resolution silver-halide material was used instead of photopolymer film. The recording has to be done at a relatively large distance from the Lippmann Optical Variable Device (OVD) and thus, it is difficult to obtain a high-resolution image of the Lippmann recording. The Lippmann photograph itself has an extremely high resolution provided that a high quality photographic macro-lens is used in the Lippmann recording equipment. The recording film has a resolving power of about 10,000 lines/mm.

Figure 4.9: Lippmann Photograph (Lippmann OVD) of the passport page

As seen in Figure 4.9, all the passport information details are recorded with high resolution in the Lippmann OVD.

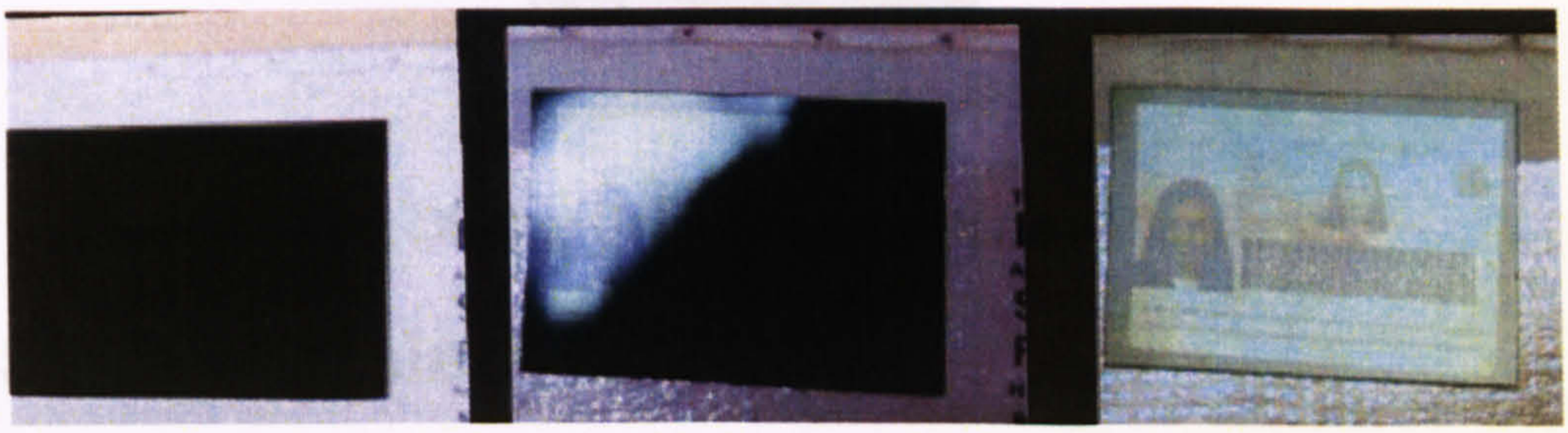
In Figure 4.11, the OVD was illuminated and recorded perpendicularly, which means that a correct colour image of the passport is possible to obtain. The colour image changes towards the blue part of the spectrum when the image is studied under oblique illumination and observation angles.

Figure 4.10: Passport page with Lippmann OVD when the OVD is not visible

Figure 4.11: Passport page with the visible Lippmann OVD when illuminated with perpendicular diffuse light and perpendicular observation

A thin Lippmann OVD film label (size: 17mm x 24mm) was laminated to the passport page in the upper left hand corner. The small Lippmann OVD label is backed with black plastic foil. In Figure 4.12(a), the label looks completely black since no diffuse perpendicular light was hitting it, nor was it observed perpendicularly. In Figure 4.12(b), the passport page was slightly tilted and in Figure 4.12(c), the whole Lippmann OVD is visible. At this angle, the passport page is perpendicular to the observer. The glossy

over laminate reflects a lot of light towards the eye and the reproduction camera lens. Upon inspecting the passport, this effect is easily and immediately observed and very difficult to simulate in any other way. Further, the Lippmann OVD image is an exact copy of the passport page itself. The only tool needed for this is a magnifying glass in order to be able to see and read the text in the image.



OVD image not visible (a) OVD image partly visible (b) Entire OVD image visible (c)

Figure 4.12: The effects of tilting the Lippmann OVD attached to the passport page

Display Unit for the Lippmann Photographs: A display device based on a Visual Plus Slide Illuminator turned upside down, mounted in a holder, was made to view the Lippmann photographs. This provides uniform white light and could enable, e.g. the passport officer to check the documents with ease to detect any fraud. Figure 4.13 demonstrates the use of the display unit.

Figure 4.13: Display Unit for the Lippmann Photographs

4.5 Conclusions

The Lippmann photographic technique can be optimized further by taking the following points into consideration.

- Exact solutions to the multiple scattering problems can be found when the angular distribution of the light scattered by the individual particles is of the simple form, for example isotropic. Generally however, this is not the situation, and approximate methods of solutions have to be developed. The goal however, is that the approximations made to provide a mathematical solution will still provide the desired accuracy to extract meaningful information.

- A more realistic models of the emulsion can be developed to include effects such as multiple scattering and absorption and with a comparison with practical results aim to improve the existing models which assume a perfect emulsion with relatively no scattering. The sensitometric and imaging properties of an emulsion are principally controlled by using various methods of halide precipitation and spectral and chem-

ical sensitisation, since these procedures almost totally determine the light-scattering and absorption properties of the grains. Methods to improve the sensitivity and image sharpness of the emulsion can be developed, by using techniques such as coating the emulsion by a reflective support which usually increases the speed of the emulsion though with a loss in sharpness or by adding light-scattering particles directly to the emulsion. The image sharpness can be improved by adding intergrain absorbing dyes or particles to the gelatin to decrease the amount of light scattered sidewise by the grains.

- The development process can be fine tuned to get more appreciable results for the end product to be commercially viable. With the film samples obtained, a spectrophotometer can be used to measure the amount of light reflected or transmitted at each wavelength of the spectrum. This method is widely used in industry and gives more detailed information about the light-reflecting or transmitting properties of the sample. Here, a spectrophotometric curve is plotted showing the reflectance or transmittance (or density) at each wavelength. These spectrophotometric curves do not directly specify the colour of a sample but calculations can be made with the aid of the CIE chromaticity diagram[101] to obtain a set of three numbers which are analogous to the red, green and blue measurements made by the filter method.

- During the development process, another important factor that needs to be taken into consideration is the time of development. Two important factors: fog and graininess related to the emulsion used increase with the time of development. Fog is the term applied to the density, which is obtained on parts of the sensitive surface, which have received no exposure. Graininess on the other hand is dependent upon the size of the developed grains and on the manner in which they are clumped

together.

Thus, the Lippman OVD has been successfully used as a security device for passports. The 19th Century technique has been practically demonstrated using Slavich photographic plates & film and the Lumiere & GP8 developers. The next part of the thesis moves on to Texture Watermarking and Microbar Covert Bar Coding Technique.

Chapter 5

Covert Bar Coding

5.1 Watermarks in Context

A digital watermark is a digital signal or pattern inserted into a digital "document" (e.g. text, graphics, multimedia presentations). As such, it is a form of electronic watermark much like the corporate logos used by the cable television industry to identify the source of the program on screen, typically along the lower periphery of the television screen. Such cable companies, we may assume, feel that the advertising advantage of the ever-present, on-screen logo, together with the legal benefit of having a source signature persist under video recording, more than offset the aggregate user-annoyance and distraction.

Digital watermarks extend these advantages to digital documents. A signal or pattern may be digitally imposed on a document prior to sale or distribution. The persistence of the watermark under transmission, and some common forms of transformation, contribute to our ability to authenticate copies. This, in turn, should enable us to protect our ownership rights in digital information, even in the undisciplined, anarchistic world of the Internet

(see Figure 5.1).

Figure 5.1: Digitized copy of artwork from a sixteenth century Aztec manuscript. Note that the circular digital watermark is most visible against light background. Faint watermarks tend to "hide" in the intense, foreground imagery[Source: IBM's Digital Library Project. Used with permission.]

In order to explain the Digital Watermarking concept, it is much easier to explain what it is not! Digital watermarking is not encryption, which also involves file transformation. It is a common practice nowadays to encrypt

digital documents so that they become un-viewable without a decryption key. Unlike encryption, however, digital watermarking leaves the original image or (or file) basically intact and recognizable. Further, decrypted documents are free of any residual effects of encryption, whereas visible digital watermarks are designed to be persistent in viewing, printing, or subsequent re-transmission or dissemination.

Digital watermarking is also to be contrasted with digital fingerprinting, which produces a "meta" file that describes the contents of the source file. Cyclic redundancy checking and checksum algorithms are both simple uses of file fingerprinting for error detection applications. A more advanced use of fingerprinting is to be found in RSA Data Security's use of message digests for authentication purposes. Digests are the result of applying a hashing algorithm (e.g., MD5, SHA) to a document or file to produce an identifying bit string (fingerprint). If the receiver's hash algorithm produces the same message digest for the file as the sender's, the file is authentic. Of course, this assumes that sender and receiver use the same software, hence the same hash algorithm.

Fingerprints may also serve as a digital signatures. If the message digest discussed above were further encrypted, converted to plaintext, and attached to the original file or message in transit, the plaintext version of the message digest (fingerprint) would also serve as a digital signature for the original file. While both fingerprints and signatures accompany unaltered source documents, signatures, like their penned counterparts, are embedded in the document itself even if in encrypted form.

5.2 Watermarks in Use

Authentication is but one use of digital watermarking. Both symmetric and asymmetric hashing algorithms can be used to embed a unique digital imprint on a document or file. If the removal of an imprint yields the original document (which is to say that the "stripped" watermark is identical to the embedded watermark), then the copy is authentic. Once again, this assumes that the "stripping" algorithm is available to the end-user. Such authentication techniques are usually associated with some sort of encryption for the distribution of keys, programs, etc, which are related to the watermarked documents.

In addition, watermarks are also used as a check for non-repudiable duplication and transmission. In this case, the owner, creator or sender imprints a watermark which is unique for each receiver. The watermark holds under subsequent re-transmission, so the "authorized" source of unauthorized copies may be easily identified after extraction. A collateral benefit is that the intended recipient of a document token could always be identified.

However, these applications really only apply to the class of invisible watermarks. Visible watermarks contribute to document and transmission security in different ways. To illustrate, visible watermarks are more overt means of discouraging theft and unauthorized use both by reducing the commercial value of a document and making it obvious to the criminally inclined that the document's ownership has been definitively established. We observe that invisible watermarks only have this effect if the digital thief is aware of the technology and the possibility that watermarks may be present on a document of interest.

There are several characteristics of effective watermarks. For one, they must be difficult or impossible to remove. For another, they must survive

common document modifications and transformations (e.g. cropping and compressing image files). Third, they must, in principle at least, be easily detectable and removable by authorized users with such privileges (e.g. law enforcement agencies). Invisible watermarks should also be imperceptible, while visible watermarks should be perceptible enough to discourage theft but not perceptible enough to decrease the utility or appreciation of the document.

5.3 Watermarking Practice

Watermarking techniques tend to divide into two categories, text and image, according to the type of document to be watermarked. In the case of imagery, several different methods enable watermarking in the spatial domain from simply flipping low-order bits of selected pixels to superimposing watermark symbols over an area of a graphic. Spatial domain watermarking is illustrated in Figures 5.2(a) and 5.2(b) that demonstrate how the degree of visibility of the watermark depends upon its intensity and the nature of the background.

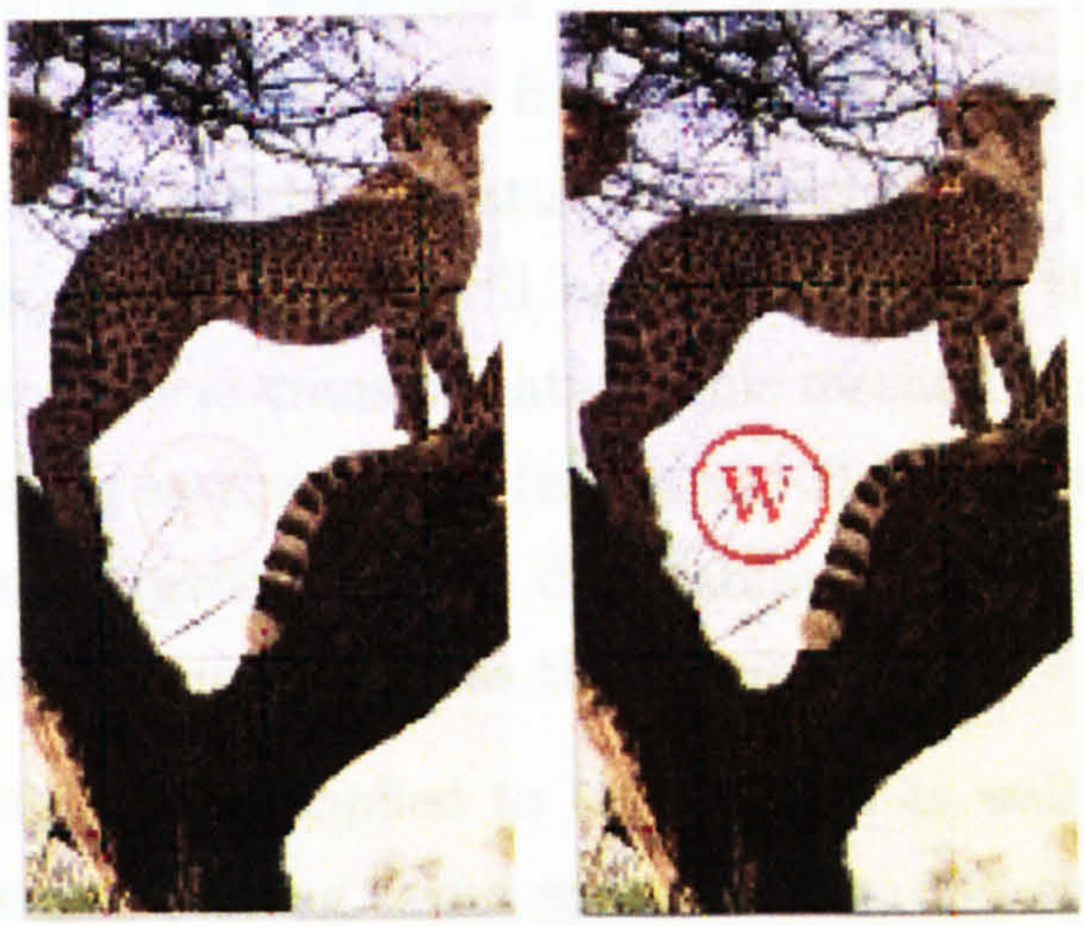


Figure 5.2: Two watermarked images identical but for the intensity of the image. Considerable latitude is available, in terms of placement, size and intensity to blend the watermark into a graphic[Source: IBM's Digital Library Project. Used with permission.]

Another spatial watermarking technique uses color separation. In this way, the watermark appears in only one of the color bands. This renders the watermark visibly subtle such that it is difficult to detect under regular viewing. However, the watermark appears immediately when the colors are separated for printing. This renders the document useless to the printer unless the watermark can be removed from the color band. This approach is used commercially for journalists to inspect digital pictures from a photo-stockhouse before buying un-watermarked versions.

An alternative to spatial watermarking is frequency domain. In this case, transforms like the Fast Fourier Transform (FFT) alter the pixel-values of the image for chosen frequencies. Since high frequencies will be lost by

compression or scaling, the watermark signal is applied to lower frequencies, or better yet, applied adaptively to frequencies that contain important information of the original picture (feature-based schemes). Since watermarks applied to the frequency domain will be dispersed over the entirety of the spatial image upon inverse transformation, this method is not as susceptible to defeat by cropping as the spatial technique. However, there is more of a tradeoff here between invisibility and decodability, since the watermark is in effect applied indiscriminately across the spatial image.

Watermarking can be applied to text images as well. Three proposed methods are: text line coding, word space coding, and character encoding. For text line coding, the text lines of a document page are shifted imperceptibly up or down. For a 40-line text page, for instance, this yields 240 possible codewords. Figure 5.3 illustrates text line coding as it would appear to the casual reader. According to the line code box, the first, second, fourth and sixth lines are elevated by 1 pixel, although the alteration is practically imperceptible. The effectiveness of such watermarking is confirmed in Figure 5.4. Even with the affected lines set apart in red, it is still difficult to determine that the lines are elevated.

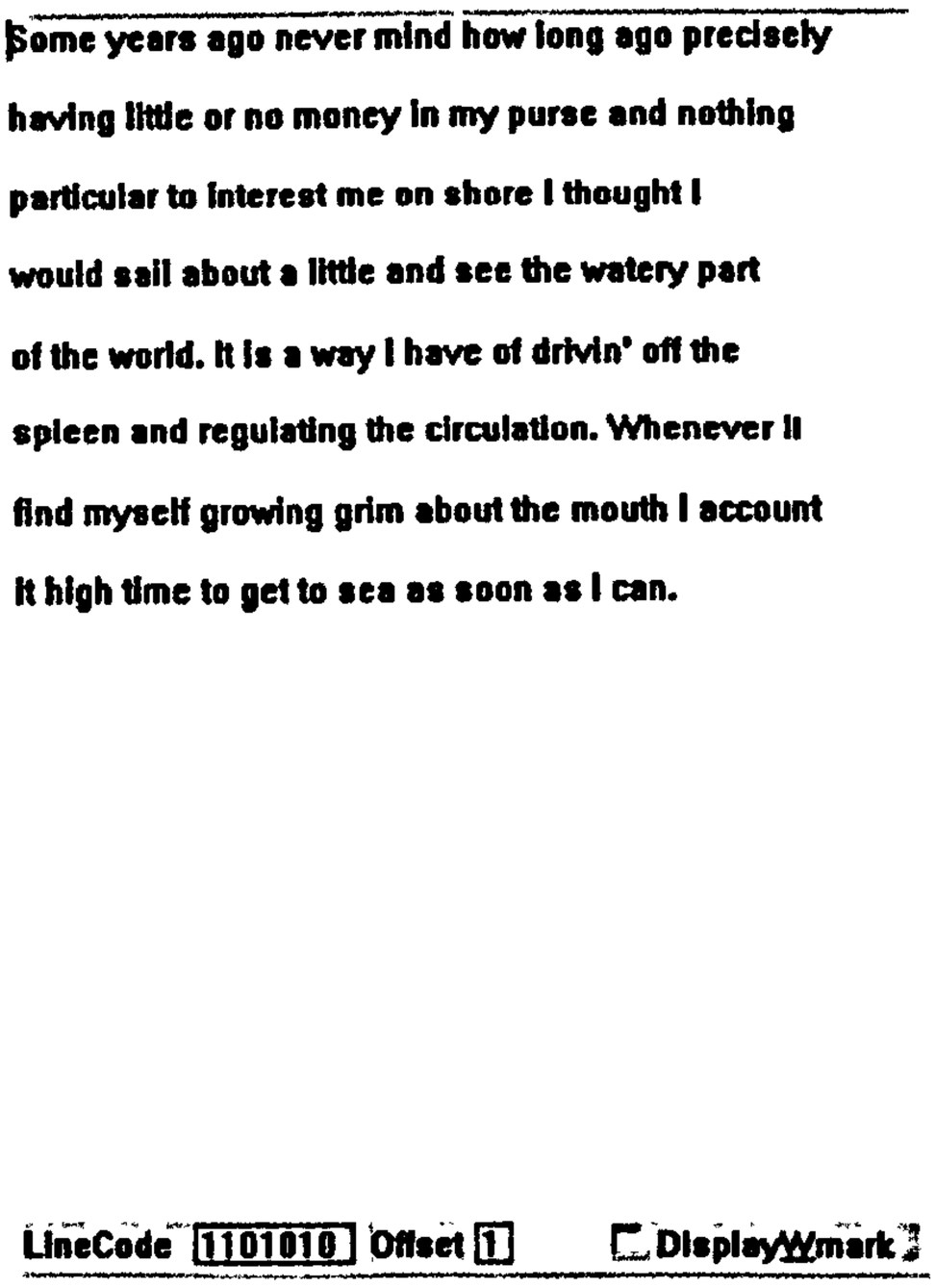


Figure 5.3: Text with lines 1,2,4 and 6 elevated from normal position by one pixel[Source: IBM's Digital Library Project. Used with permission.]

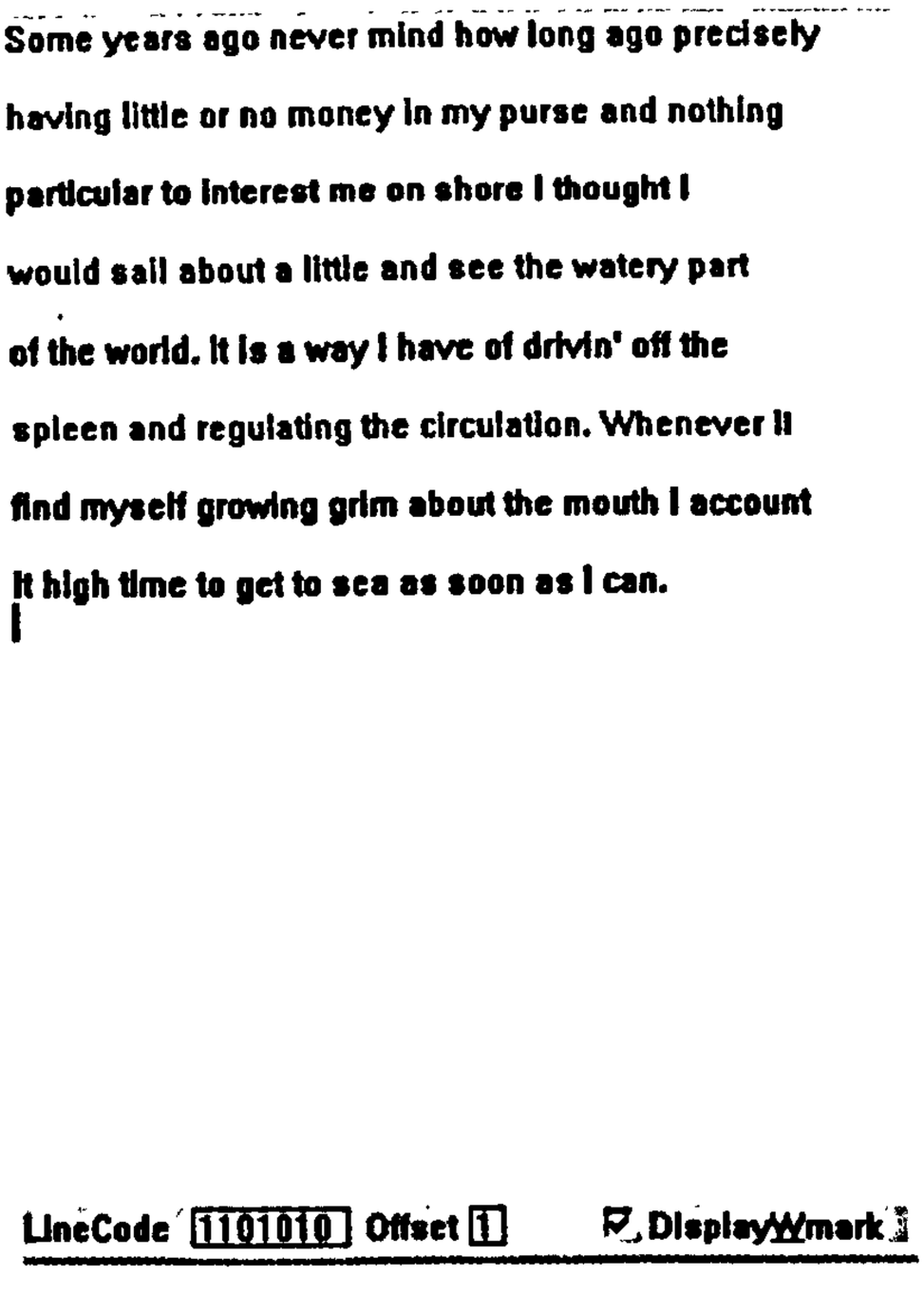


Figure 5.4: Elevated lines highlighted[Source: IBM's Digital Library Project. Used with permission.]

For word-shift coding, the spacing between words in a line of justified text is altered. The plaintext in Figure 5.5 has three words shifted right one pixel. Figure 5.6 highlights the affected words.

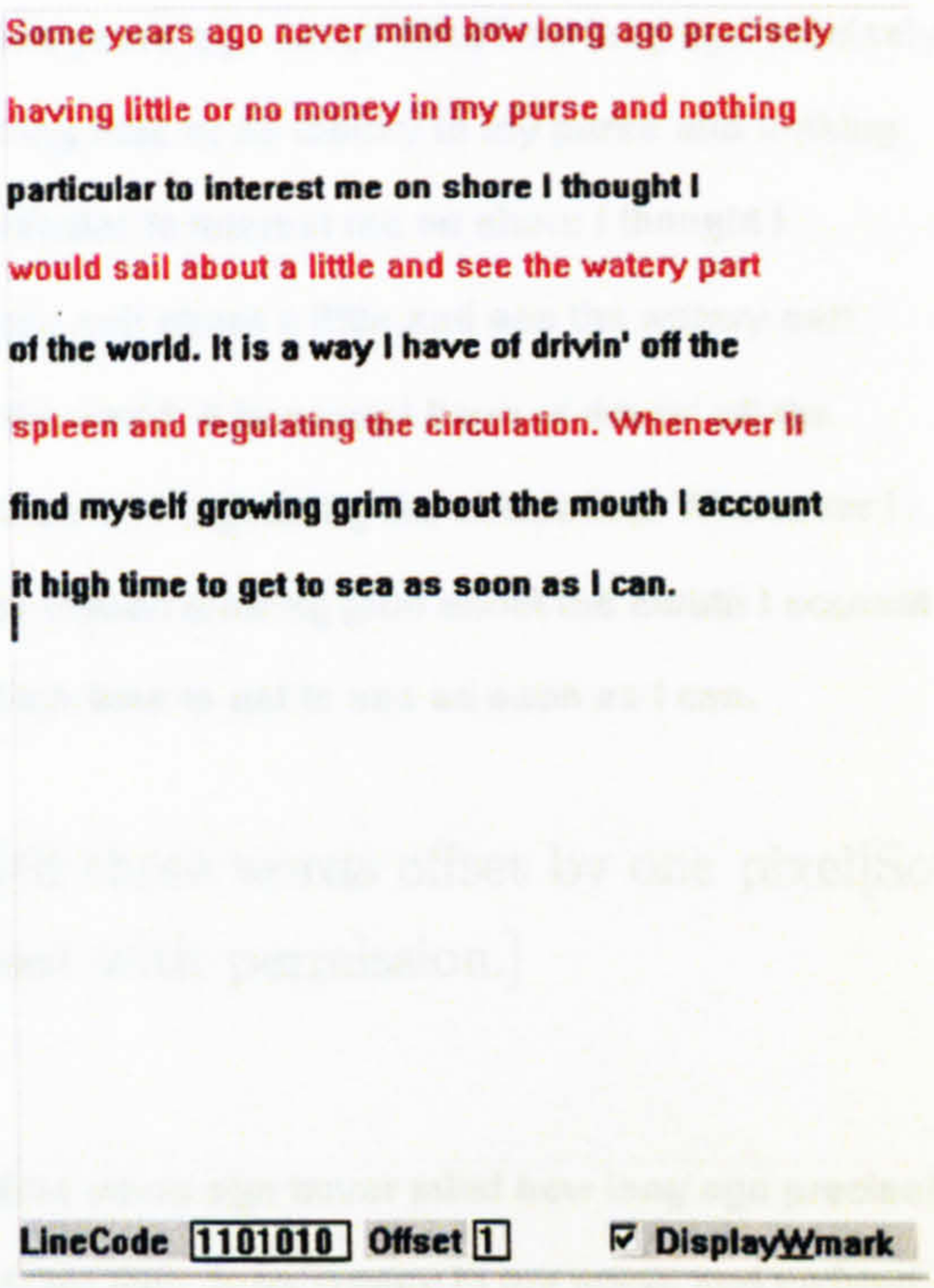


Figure 5.4: Elevated lines highlighted[Source: IBM's Digital Library Project. Used with permission.]

For word-shift coding, the spacing between words in a line of justified text is altered. The plaintext in Figure 5.5 has three words shifted right one pixel. Figure 5.6 highlights the affected words.

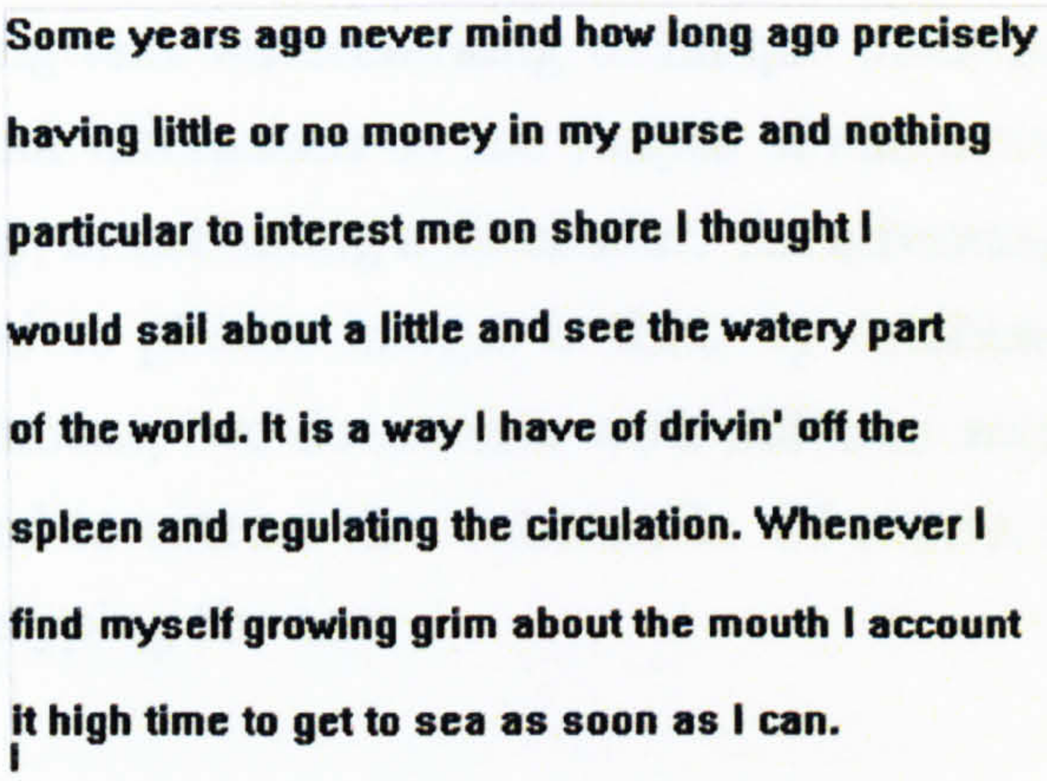


Figure 5.5: Text with three words offset by one pixel[Source: IBM's Digital Library Project. Used with permission.]

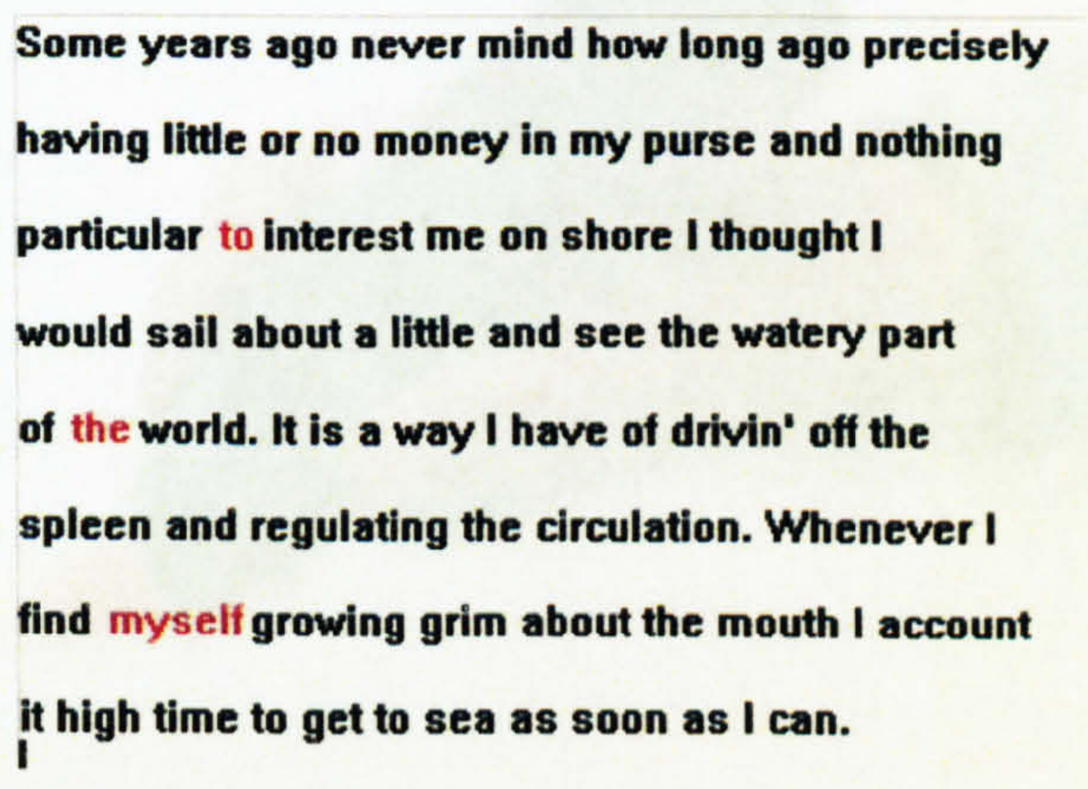


Figure 5.6: Text with offset words highlighted[Source: IBM's Digital Library Project. Used with permission.]

The remaining text watermarking technique involves character coding. This involves minor alterations to the shapes of characters - e.g. clipping a serif imperceptibly, or extending a descender. An advantage of these methods over those applied to picture images is that, by combining two or three of these to one document, two documents with different watermarks cannot be spatially registered to extract the watermark. Of course, the watermark can be defeated by retyping the text.

5.4 Microbar Pen Scanner

Figure 5.7: The Microbar Pen Scanner [Adapted from Quick Link Pen Operation Manual]

Human interpretation of an image is based on matching templates of geometrically significant features. Microbar uses fractal geometry to encode data in the background of a document i.e. fractals are used to create appropriate image texture to camouflage information. This technique is unique and has

widespread commercial applications. To mention a few of its advantages; the system offers a higher degree of security than any other encryption method known; the technology used to implement it is very standard and simple and therefore can be used for a wide range of products. The main feature is its covert nature, which makes it invisible to the naked eye and hence difficult for the counterfeiter to imitate.

Equipment used: Desktop computer with a standard Epson 1240U scanner and also a hand held scanner called the Quick Link Pen which is more convenient to use and it can store, edit, transfer text for easy management. It can also synchronize and transfer data to desktop applications in seconds.

Normally, in image processing, we like to have an image with as much less noise as possible, as noise increases the distortion and reduces the resolution of the image. The aim of the Microbar, however, is to introduce noise in the image or document which is to be watermarked and the noise generated through fractals serves as the watermark and has the covert feature. This is difficult for counterfeiters to regenerate as the exact parameters involved in the generation of the fractal noise (watermark) is unknown except to the person embedding the watermark. Thus the idea is novel and unique and has generated much interest for various commercial applications.

5.5 Covert Bar Coding: *The Idea*

- A new product (invisible signature embedded into images) designed to work alongside Microbar authentication.
- Facilitates track & trace down to product individualisation level.
- Can be used with Digital Printing to allow local control.

- Completely invisible for fractal images (e.g. clouds, see Figure 5.44), encrypted and covert bar coding.
- Simultaneous authentication (anti-counterfeit detection) & product individualisation reading by scan.
- Machine readable using flat bed or pen scanner.
- Cannot be copied or reverse engineered.

So to summarise, the two fundamentals underpinning Microbar are Fractals and Chaos. Fractals give the characteristic of self-similarity and the notion of 'texture'. This combination gives Microbar both its ability for partial scanning and scanning in any direction as well as its inherent camouflaging potential for images with intrinsic 'texture' or complexity and detail. It allows the scanning at much lower resolutions than those used for printing of the document being scanned.

Chaos is an enabling technology which is used to generate self-similiarity. Intuitively, people find the concept of structure from chaos difficult to grasp. The Faigenbaum map in the figure below explains this in more detail. Chaos also gives an extremely powerful encryption capability. Therefore, this technology is used in a stand-alone encryption product. Thus, this makes the Microbar even more secure.

The map in Figure 5.8 is a way of showing the statistical self-affinity generated by chaos. The Faigenbaum map plots the output from a chaotic algorithm which requires a 'seed' or parameter to drive it. In the above example, the various values of this 'seed' are shown on the horizontal axis. At first, for values upto 2.0, the algorithm outputs converge to a single number -1.0. As the 'seed' value goes above 2.0, a phenomenon known as bifurcation occurs and one of two separate numbers are repeatedly generated. At

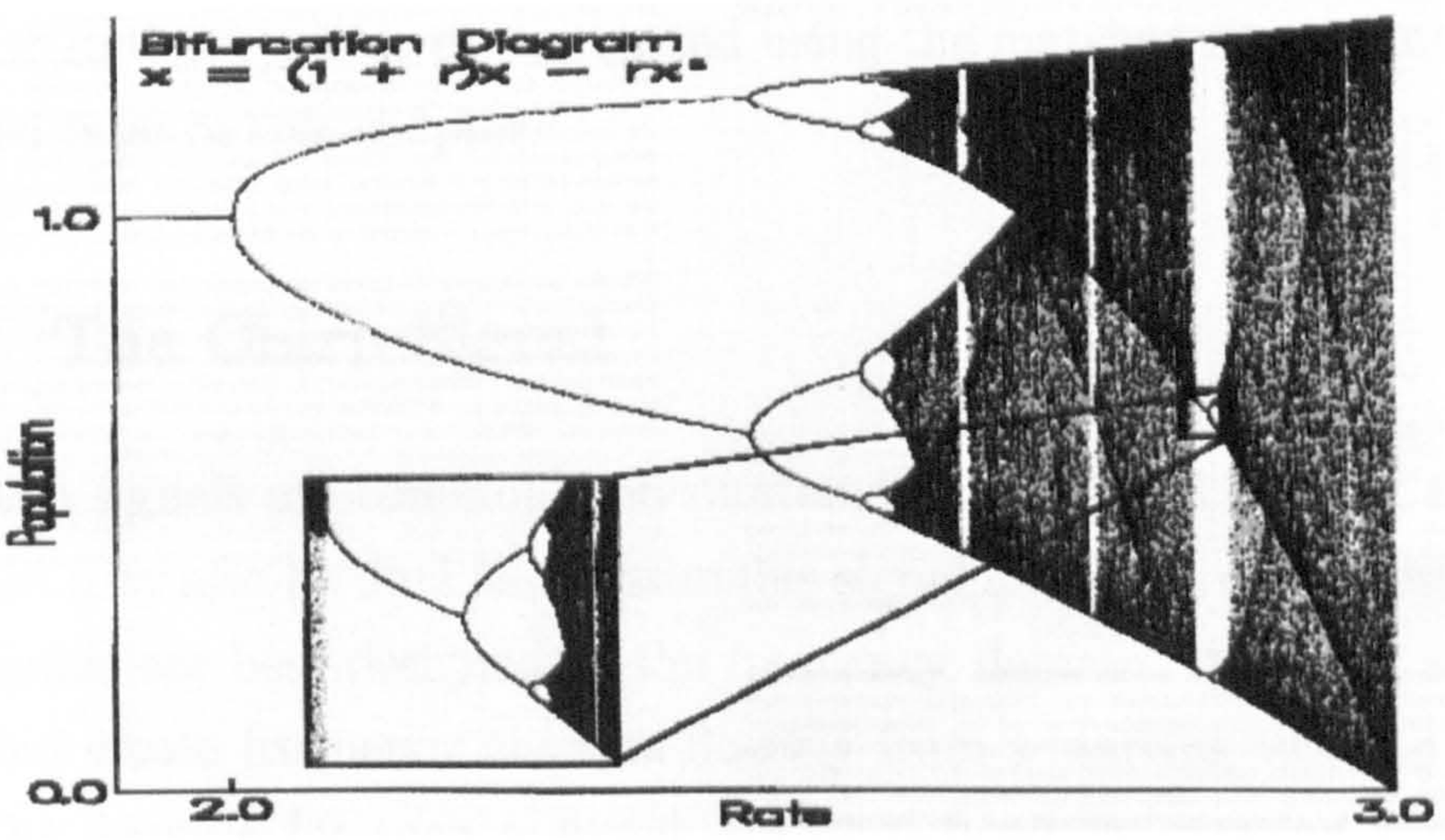


Figure 5.8: The Faigenbaum Map[111]

a 'seed' value of around 2.5, bifurcation occurs again with one of four possible numbers being generated. At 2.6, bifurcation occurs again generating 8 possible answers. Beyond this point, we start entering the black shaded area which represents mathematical chaos, where any number within the limits is able to be output by the algorithm. The interesting points to note are the 'white bands' within the black 'sea of chaos'. On closer inspection, these bands represent structures similar to the previous points of bifurcation. This is mathematical self-similarity which links chaos to fractals.

5.6 Covert Bar Code: *The Fundamentals*

5.6.1 Introduction

The 'Microbar Covert Bar Code' is fundamentally based on the use of the 'chirp' function. The 'chirp' signal is then convolved with a bar code and then added to an image for watermarking. It is covert for low signal-to-noise

ratios (SNRs) and can be reconstructed using the matched filter principle as presented later in this chapter.

5.6.2 The Chirp Signal

Chirp signals are commonly encountered in radar and ranging applications, but may also be used as the stimulus signal for device characterization. Such signals are best designed in the frequency domain. A 'chirp' signal is a sinusoid whose frequency changes linearly from a starting value to ending one. The formula for such a signal can be defined by creating a complex exponential signal with quadratic phase which is given by

$$p(\tau) = \exp(ik_0\tau)\exp(i\alpha\tau^2)\exp(\phi), \quad -T/2 \leq \tau \leq T/2$$

where T is signal length, τ is time $\times$ speed of light, α is quadratic chirp rate/(speed of light)^2, k_0 is the carrier wave number (carrier frequency = $\frac{k_0}{2\pi} \times$ speed of light) and ϕ is a phase value.

In reality, the signal is not complex, but a real valued function of time. Hence, the real part of p is given by

$$\cos(k_0\tau + \alpha\tau^2 + \phi)$$

The instantaneous phase of this signal is $\varphi = k_0\tau + \alpha\tau^2 + \phi$. The derivative of φ yields an instantaneous frequency modulation of $f_i(t) = k_0 + 2\alpha\tau$ which is linear versus τ. Hence, the reason why it is also known as the linear frequency modulated (FM) signal.

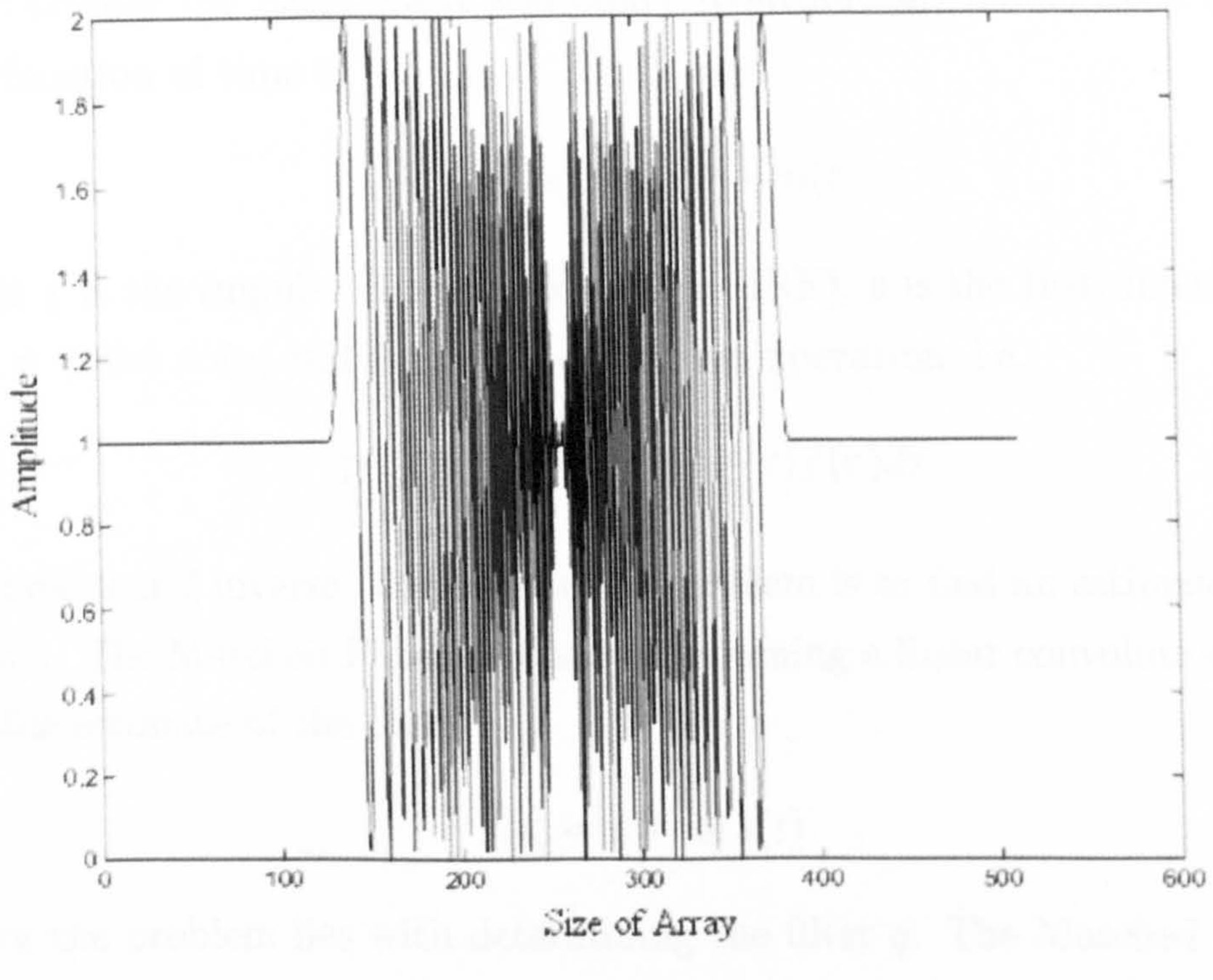


Figure 5.9: An example of a chirp signal

5.6.3 The Matched Filter

The Matched Filter is one of the most common filters used for pattern recognition. It uses the principle of correlating a signal/image with a matching template of the feature that is assumed to be present in the signal/image. If the feature exists, then the filter output i.e. the correlation signal/image, produces a local maximum or spike where the feature occurs. This is a general process but has special significance when the template and feature are based on the chirp function as it produces a robust output when

the signal-to-noise ratio is very low.

Consider the basic linear stationary (convolution) model for a signal s as a function of time t,

$$s(t) = p(t) \otimes f(t) + n(t)$$

where f is the Impulse Response Function (IRF), p is the Instrument Function, n is the noise and $\otimes$ is the convolution operation, i.e.

$$p(t) \otimes f(t) = \int p(t-\tau) f(\tau) d\tau$$

A fundamental inverse (deconvolution) problem is to find an estimate $\hat{f}$ of f given s. The Matched Filter is based on assuming a linear convolution model for this estimate of the form

$$\hat{f}(t) = q(t) \otimes s(t)$$

where the problem lies with determining the filter q. The Matched Filter is based on finding q subject to the condition that

$$r = \frac{|\int Q(\omega) P(\omega) d\omega|^2}{\int |N(\omega)|^2 |Q(\omega)|^2 d\omega} \tag{5.1}$$

is a maximum where Q, P and N are the Fourier transforms of q, p and n respectively and where the Fourier transform pair is defined by

$$\begin{aligned} F(\omega) &= \int_{\infty}^{\infty} f(t) \exp(-i\omega t) dt \\ f(t) &= \frac{1}{2\pi} \int_{\infty}^{\infty} F(\omega) \exp(i\omega t) d\omega \end{aligned}$$

and r is a 'measure' of the signal-to-noise ratio. In this sense, the Matched Filter maximises the Signal-to-Noise Ratio (SNR) of the output.

Assuming that the noise n has a 'white' or uniform power spectrum, the filter Q which maximises the SNR defined by r can be shown to be given by

the simple result $Q(\omega) = P^*(\omega)$. The required solution is therefore given by

$$\hat{f}(t) = \frac{1}{2\pi} \int_{\infty}^{\infty} P^*(\omega) S(\omega) \exp(i\omega t) d\omega$$

Using the correlation theorem[111], this can be rewritten as

$$\hat{f}(t) = p(t) \odot s(t) \equiv \int_{\infty}^{\infty} p(\tau + t) s(\tau) d\tau$$

Hence, the Matched Filter is based on correlating the signal s with the instrument function p.

Derivation of the Matched Filter

The Matched Filter is essentially a 'by-product' of the 'Schwarz Inequality', which is given by

$$\left| \int_{\infty}^{\infty} Q(\omega) P(\omega) d\omega \right|^2 \leq \int_{\infty}^{\infty} |Q(\omega)|^2 d\omega \int_{\infty}^{\infty} |P(\omega)|^2 d\omega \tag{5.2}$$

Using a principle mathematical trick, $Q(\omega)P(\omega)$ can be rewritten as

$$Q(\omega)P(\omega) = |N(\omega)|Q(\omega) \times \frac{P(\omega)}{|N(\omega)|} \tag{5.3}$$

Equation 5.3 is then substituted into Equation 5.2, resulting in

$$\begin{aligned} \left| \int_{\infty}^{\infty} Q(\omega) P(\omega) d\omega \right|^2 &= \left| \int_{\infty}^{\infty} |N(\omega)| Q(\omega) \frac{P(\omega)}{|N(\omega)|} d\omega \right|^2 \\ &\leq \int_{\infty}^{\infty} |N(\omega)|^2 |Q(\omega)|^2 d\omega \int \frac{|P(\omega)|^2}{|N(\omega)|^2} d\omega \end{aligned} \tag{5.4}$$

From the result in Equation 5.4 and using the definition of r in Equation 5.1, the result can be observed as

$$r \leq \int \frac{|P(\omega)|^2}{|N(\omega)|^2} d\omega$$

If r is to be maximum, then the required result is

$$r = \int \frac{|P(\omega)|^2}{|N(\omega)|^2} d\omega$$

or

$$\left| \int_{\infty}^{\infty} |N(\omega)| Q(\omega) \frac{P(\omega)}{|N(\omega)|} d\omega \right|^2 = \int_{\infty}^{\infty} |N(\omega)|^2 |Q(\omega)|^2 d\omega \int \frac{|P(\omega)|^2}{|N(\omega)|^2} d\omega$$

But this is only true if

$$|N(\omega)| Q(\omega) = \frac{P^*(\omega)}{|N(\omega)|}$$

Hence, r is maximum when

$$Q(\omega) = \frac{P^*(\omega)}{|N(\omega)|^2}$$

The above derivation is for the one-dimensional instance. However, this can easily be extended to the discrete two-dimension instance using the i and j notations and replacing $\int$ with $\sum$. For the purpose of this research, the one and two-dimensional discrete equations are used for ease of use in computational experiments.

5.6.4 Recovering 1-D Signal using Matched Filter

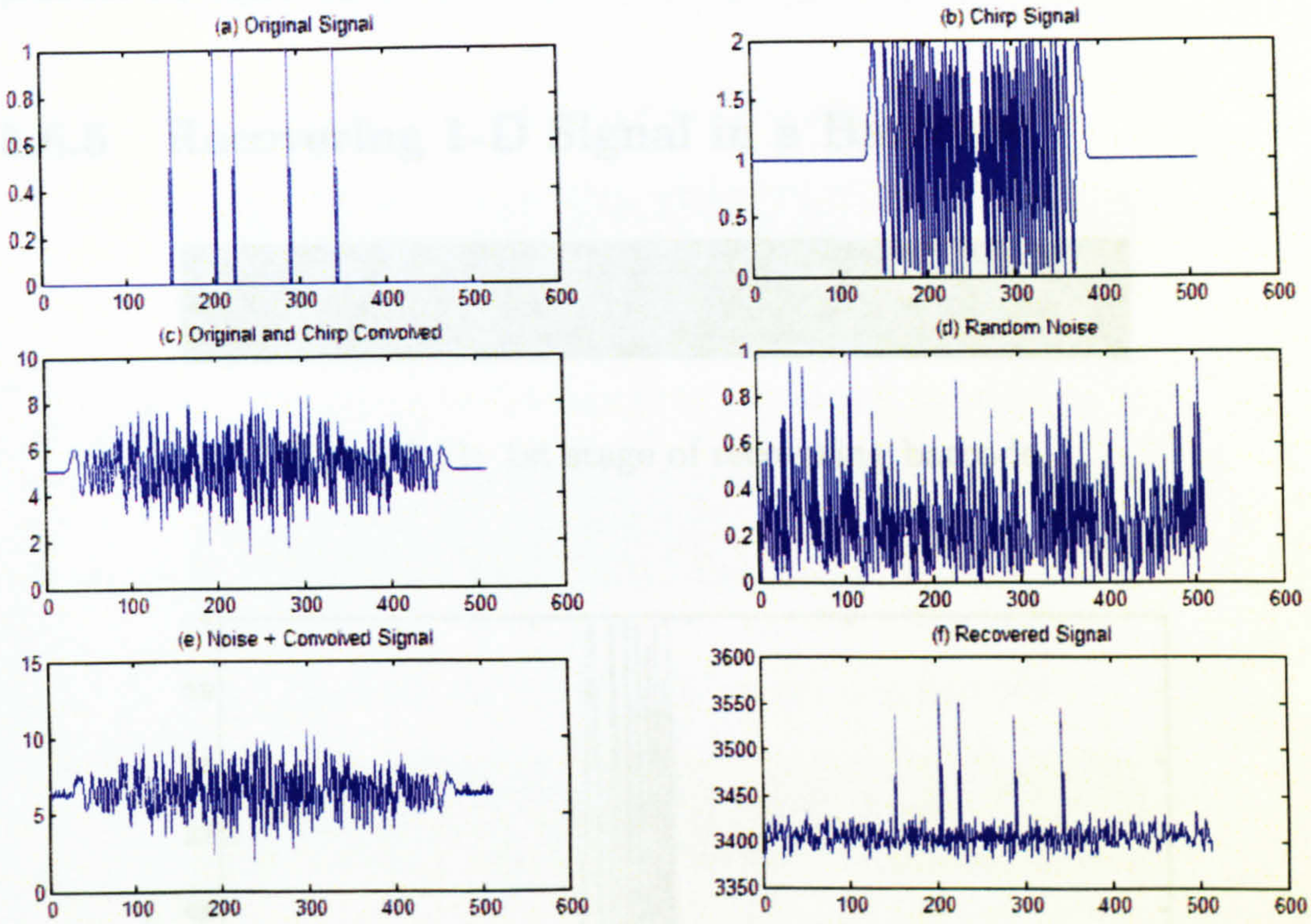


Figure 5.10: Recovery of 1-D signal using Matched Filter

With reference to Figure 5.10, (a) is a source (original) signal, (b) is the corresponding chirped pulse, (c) is a convolution of (a) and (b), (d) is a randomly generated noise signal; (e) is the addition of (c) and (d). Finally (f) is the recovered original signal using the matched filter. The original signal has to be a discreet (analog or digital) signal. We find that greater the width of the chirped pulse, greater is the resolution of the recovered signal. This principle works even when the signal-to-noise ratio is 1. The matched filter is very robust with respect to other available filters especially for high noise levels. This proves that the Matched Filter can be used to give an approximate (or exact) reproduction of the original signal.

In the following subsection, the above principle has been extended to produce an entire bar code.

5.6.5 Recovering 1-D Signal in a Barcode

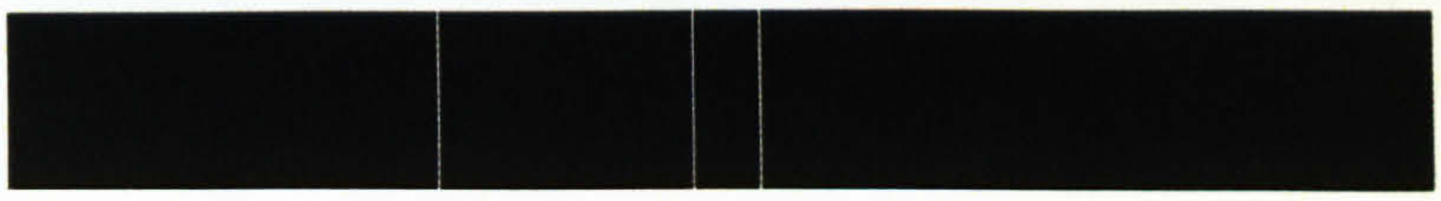


Figure 5.11: 1st stage of recovering barcode

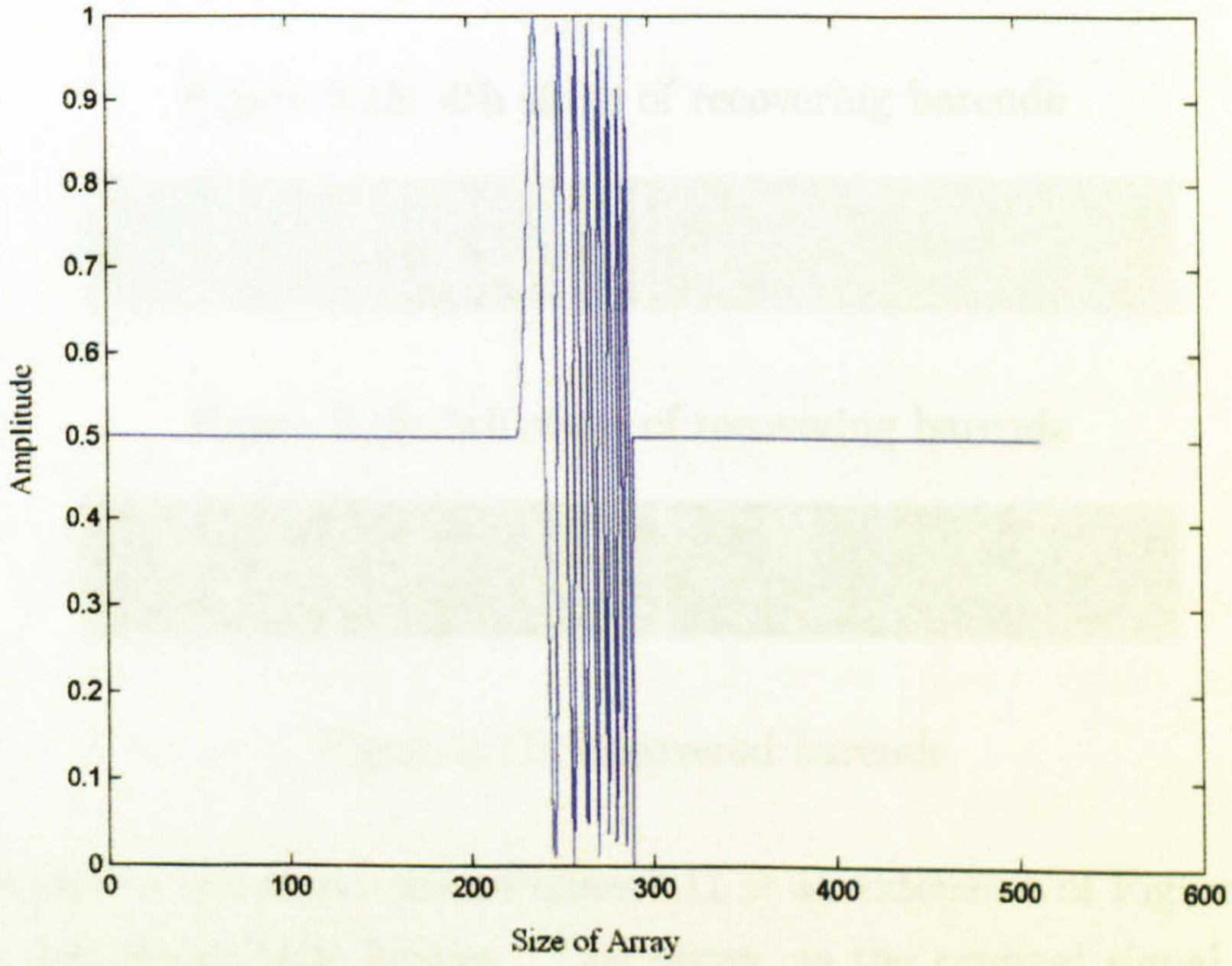


Figure 5.12: Chirp signal

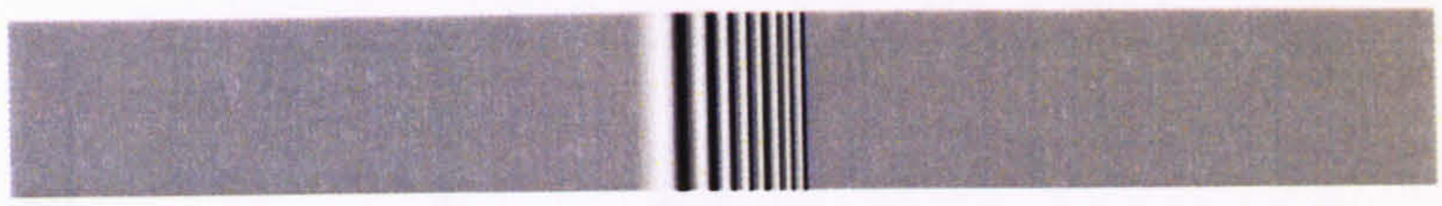


Figure 5.13: 2nd stage of recovering barcode

Figure 5.14: 3rd stage of recovering barcode

Figure 5.15: 4th stage of recovering barcode

Figure 5.16: 5th stage of recovering barcode

Figure 5.17: Recovered barcode

Similar to the above case, Figure 5.11 is an extension of Figure 5.10(a) but in two-dimensional format. This serves as the original signal. Figure 5.13 is a two-dimensional chirped pulse which has been obtained from Figure 5.12 which is a one-dimensional chirped pulse similar to the above case. The convolution of the chirped pulse with the original signal gives Figure 5.14. Randomly generated two-dimensional noise signal (Figure 5.15) is then added

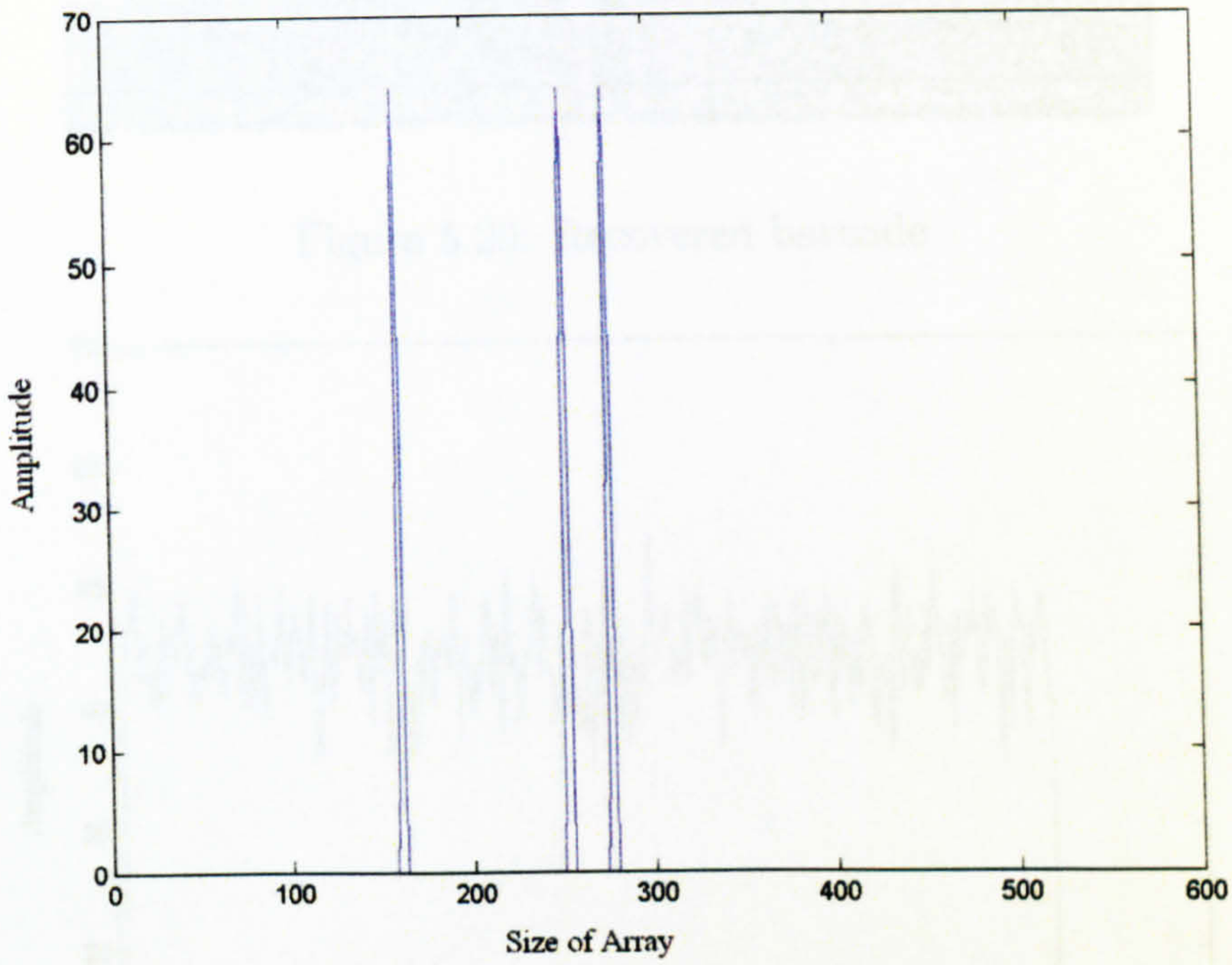


Figure 5.18: Recovered barcode signal

to Figure 5.14 (convolution of two-dimensional chirped pulse with original signal) to give Figure 5.16. The recovered barcode is then obtained in Figure 5.17 using similar principles to the above case. The actual recovered signal in one-dimensional can be seen in Figure 5.18. Thus the similarity can be seen between the original signal and the recovered signal which makes it a very good technique to be employed in watermarking.

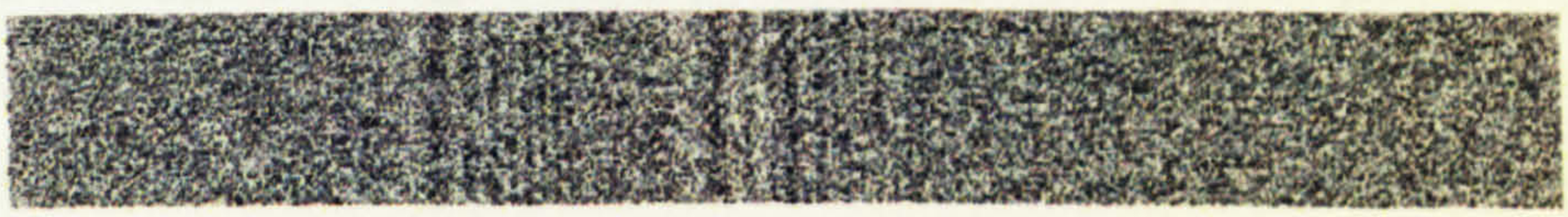


Figure 5.19: Noisy barcode

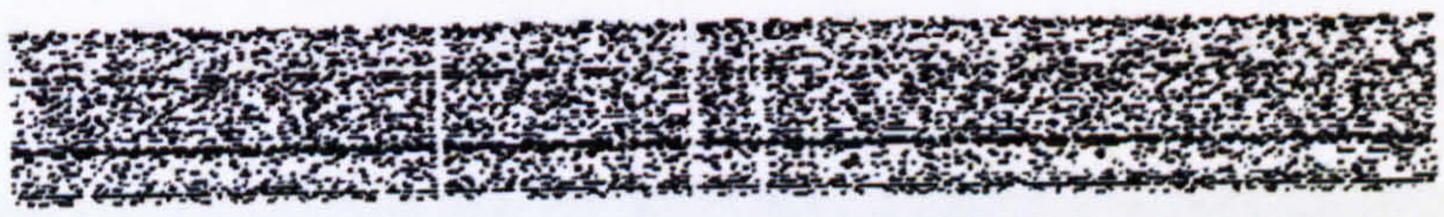


Figure 5.20: Recovered barcode

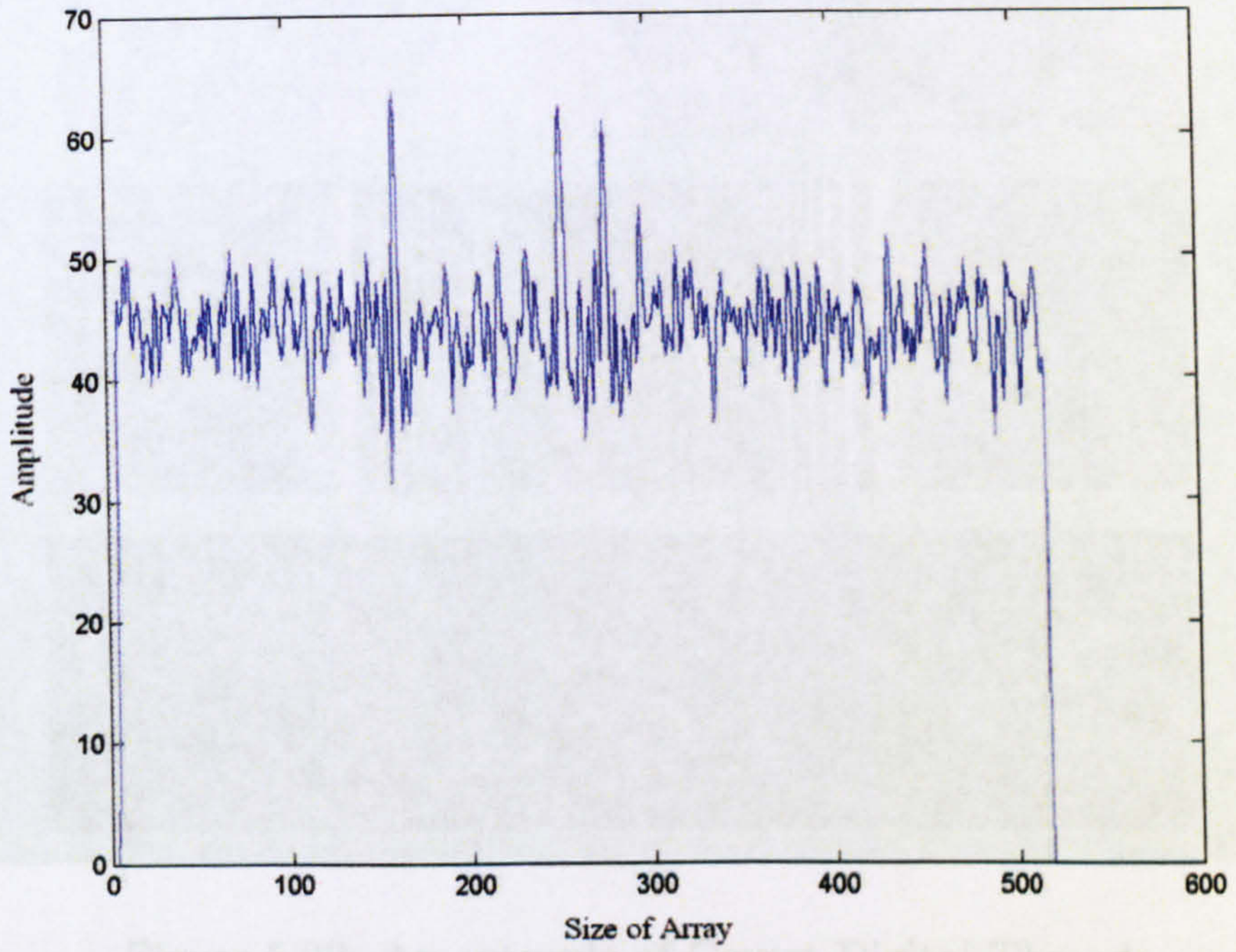


Figure 5.21: Recovered barcode signal

The randomly generated noise signal can be varied with respect to its signal-to-noise ratio and different outputs (recovered signals) can be obtained. The lower the signal-to-noise ratio (high density noise), the recovered signal remains more obscure. This is shown in Figures 5.19 to 5.21.

5.6.6 An Example of Covert Digital Thread

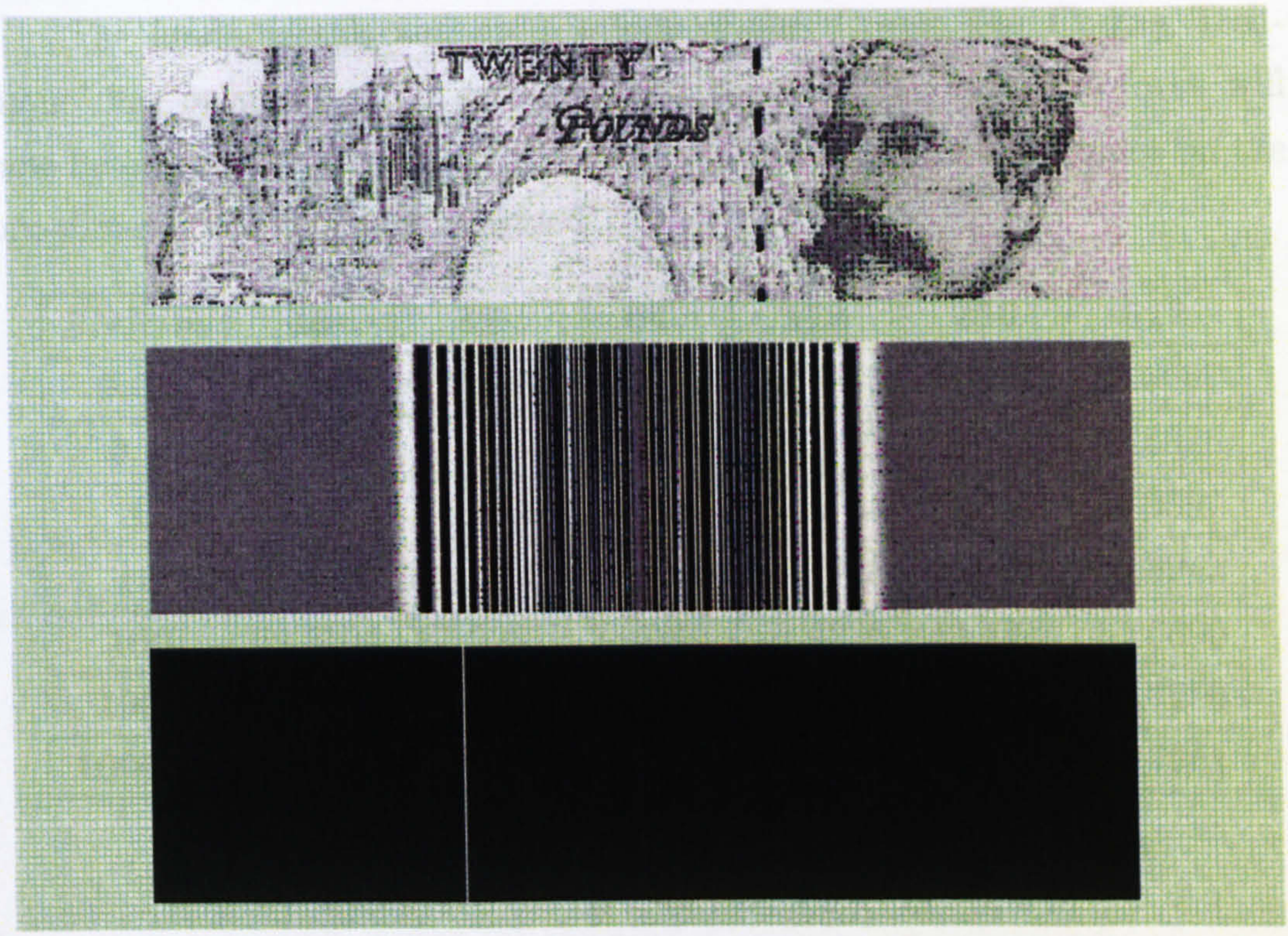


Figure 5.22: An example of Covert Digital Thread

The above figure highlights a practical application using the 'Covert Bar Coding Technique'. (Bottom-Up with reference to Figure 5.22) A single bar is encoded within the barcode. The barcode is convolved with a 'chirp' pulse and then is watermarked and embedded within the twenty pound note. By varying the opacity of the watermark for a required percentage, the twenty pound note is able to hide the watermark and is invisible to the naked eye.

5.6.7 Results and Conclusion

The fundamentals and applications of the 'Covert Bar Coding Technique' has been presented earlier in this section. However, the focus will now be on the performance of the technique. First, the performance on the one-dimensional signals is analysed.

SNR/m	16	32	64	128	256	512
0.01	92.7916	92.4282	92.6725	93.7943	93.5721	93.6557
0.1	92.7028	91.6481	92.8374	93.9329	91.4440	92.5375
1	90.1812	91.4439	88.5063	86.5754	75.0661	69.0597
10	72.2155	86.0339	77.6890	61.1428	44.6376	58.5943
100	69.5104	85.0281	78.7635	61.6226	43.4293	57.6690
1000	69.3505	85.0552	78.6805	61.1531	43.3209	57.7200
100000	69.3440	85.0527	78.6712	61.2168	43.2995	57.7351

Table 5.1: Mean Square Error percentages for 1-D Signals

With reference to the above table, SNR represents the signal-to-noise ratio from 0.01 to 100000 and m represents the width of the 'chirp' pulse from 16 to 512.

SNR can also be represented in decibel (db) units. The conversion formula is as follows:

$$SNR = 20\log_{10}(\Gamma)db$$

where Γ represents the ratio between the signal energy and noise energy. Since the above representation of SNR does not change the results shown in Figures 5.23 and 5.25, for simplicity of calculations, SNR has been represented as $\log_{10}$. However, the SNR values in Tables 5.1 and 5.2 (corresponding to Figures 5.23 and 5.25 respectively) are in decimal units.

Mean Square Error (MSE) is widely used to measure how close the reconstructed signal is to the original. The MSE gives information about the error signal energy or may be called compression noise. In terms of signal-to-noise ratio, it makes sense to compare the error energy to the original signal energy. The mean square error percentage in this case is defined by

$$\mathrm{MSE}(\%) = \frac{\sum_{i=1}^{N}(s_i - f_i)^2}{\sum_{i=1}^{N}(s_i + f_i)^2} \times 100$$

where N is the length of the signal array, i is the array index, s_i is the original signal and f_i is the recovered signal.

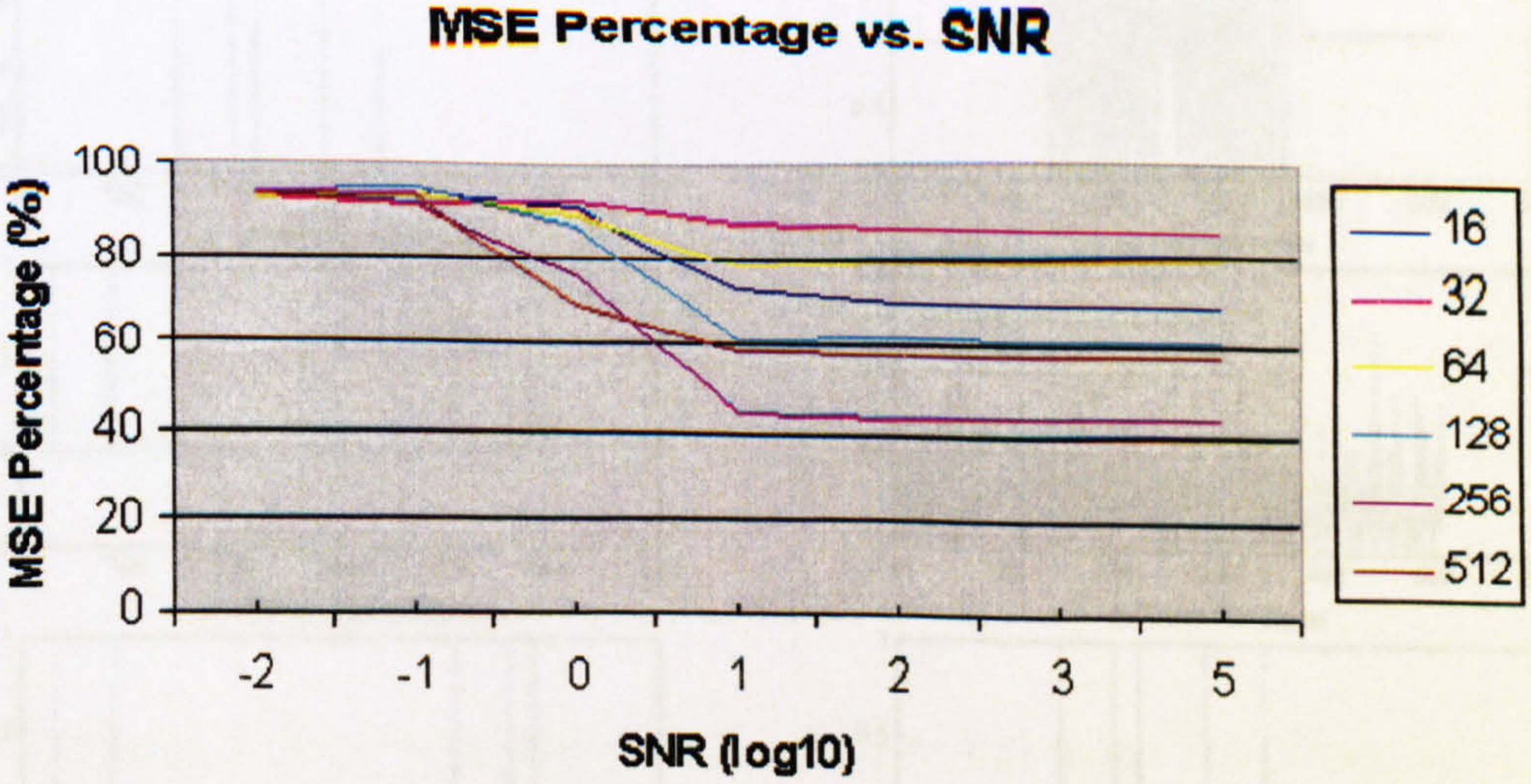


Figure 5.23: Percentage of Mean Square Error for 1-D Signals

Figure 5.23 is a graphical representation between the mean square error percentage and the log (base-10) of signal-to-noise ratio. Depending on the

width of the 'chirp' pulse (colour coded in the graph), the higher the signal-to-noise ratio, the better the recovery percentage. It can be observed that when the peak signal-to-noise ratio is reached for a particular 'chirp' pulse width, the error rate stabilizes accordingly.

However, the minimum error rate is approximately 43% (with respect to using the same sized width of the 'chirp' pulse as the original signal size), which can be considered relatively high, but on visual inspection, the recovered signal is an exact replica of the original signal accompanied by noise.

If the 'chirp' and the additional normalised noise is removed, the recovery percentage is 100%. This is depicted in Figure 5.24.

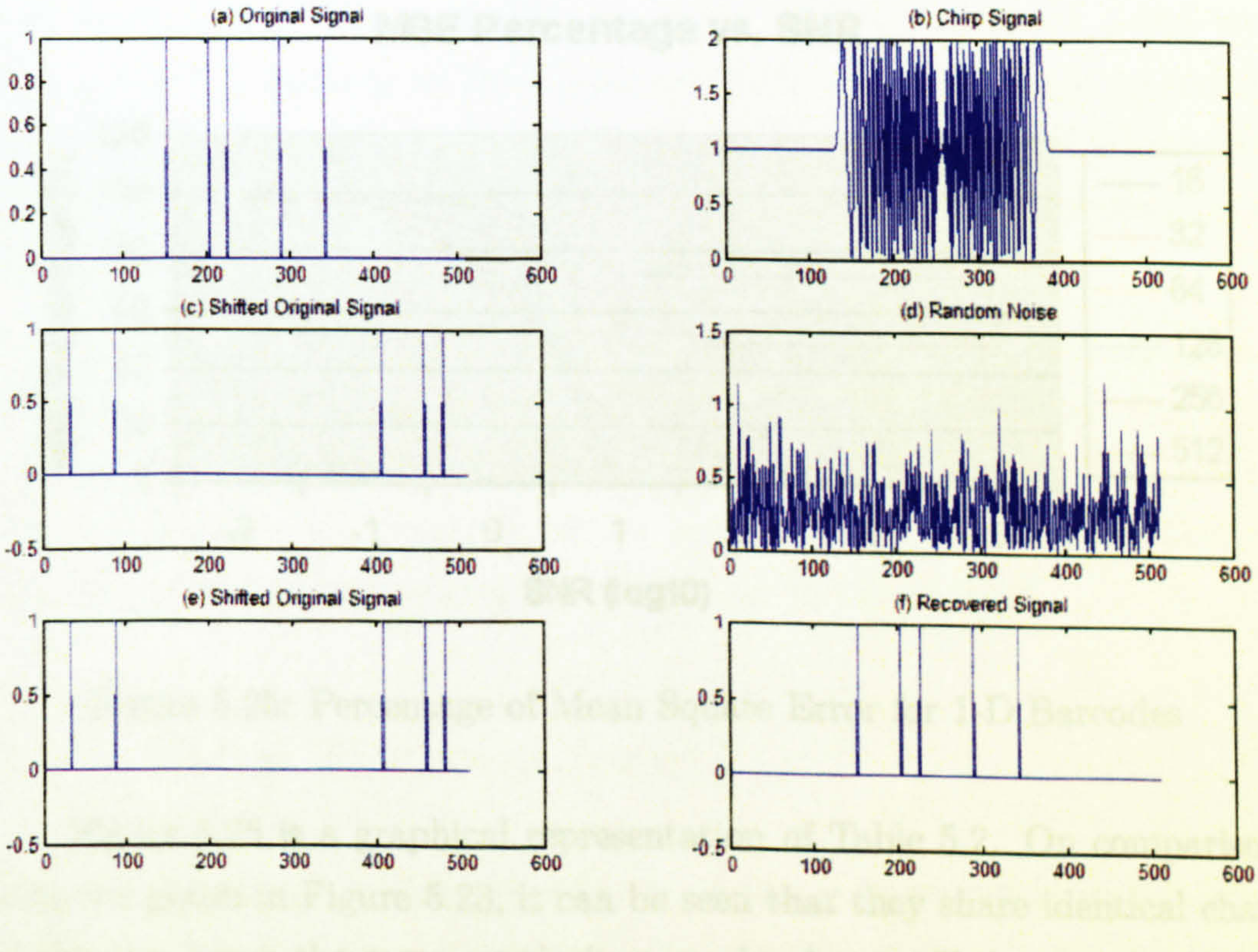


Figure 5.24: 100% recovery of signal without noise

The same principle is now applied to a one-dimensional barcode recovery.

SNR/m	**16**	**32**	**64**	**128**	**256**	**512**
0.01	97.7116	97.7832	97.8098	97.6374	97.6658	97.7056
0.1	97.4519	97.4350	97.5224	97.3662	96.9507	97.0649
1	97.0836	96.2268	96.8865	94.4764	85.8798	72.7079
10	72.5224	55.8367	33.3333	0	0	20.0000
100	71.4286	50.0000	33.3333	0	0	20.0000
1000	71.4286	50.0000	33.3333	0	0	20.0000
100000	71.4286	50.0000	33.3333	0	0	20.0000

Table 5.2: Mean Square Error percentages for 1-D Barcodes

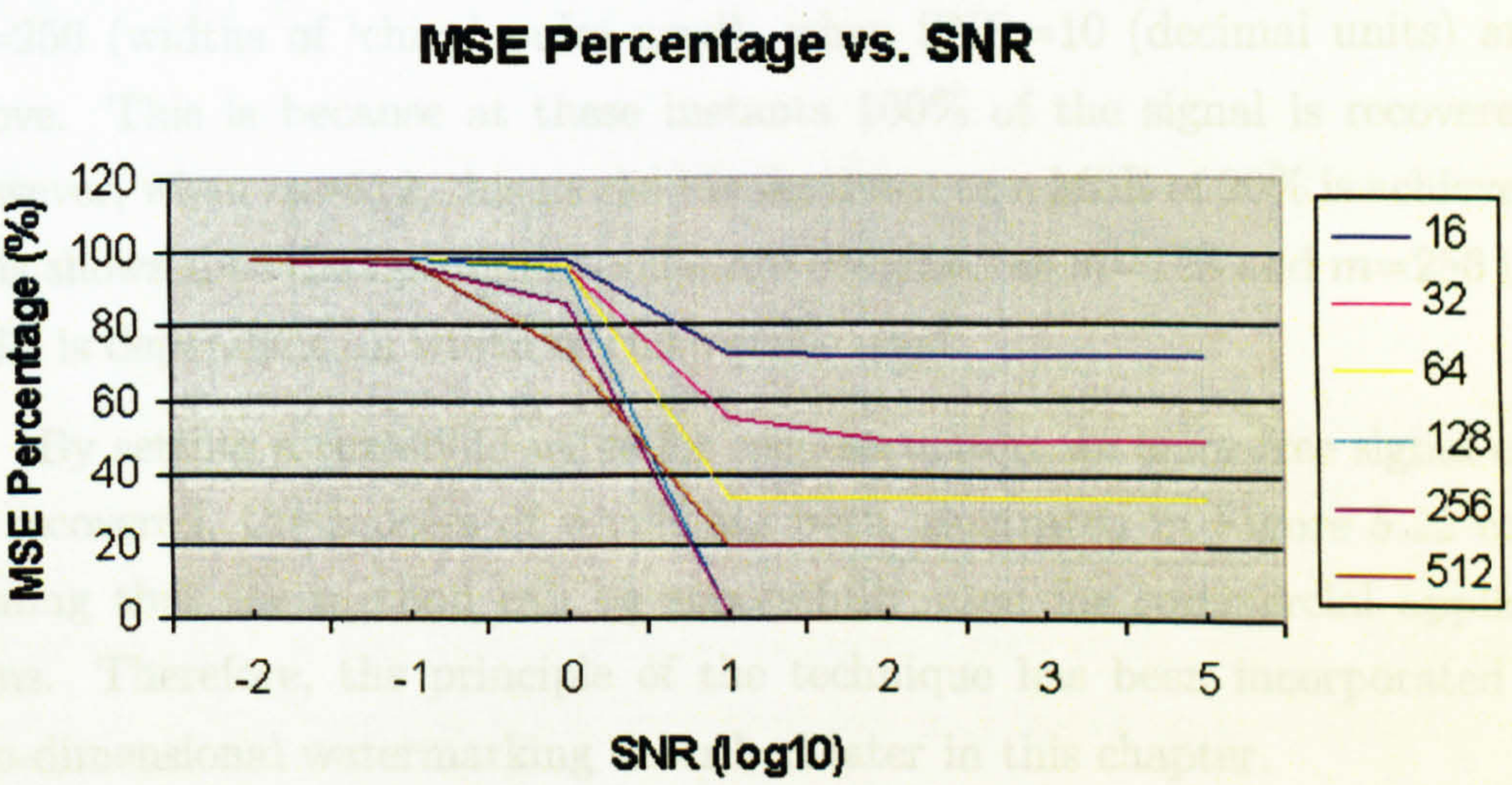


Figure 5.25: Percentage of Mean Square Error for 1-D Barcodes

Figure 5.25 is a graphical representation of Table 5.2. On comparison with the graph in Figure 5.23, it can be seen that they share identical characteristics, hence the same conclusions can be drawn. If the signal-to-noise ratio is higher, the better the recovery percentage and when the peak signal-to-noise ratio is reached (depending on the width of the 'chirp' pulse), the

error stabilizes accordingly. Peak Signal-to-Noise Ratio (PSNR) is the value of the SNR reached by the signal at stabilization, i.e. with respect to Figure 5.25, PSNR=1 or in decimal units, PSNR=10 as shown in Table 5.2. However, the difference between the two analyses is that in the first case the recovered signals are not set at a threshold and hence, accounting for the high mean square error. In the second case, the recovered barcodes are set against the threshold of its averaged noise (approximately half the difference between the maximum and minimum of the barcode intensity which is defined by the sum of all the pixel intensities within the barcode), hence the 100% recovery of the original barcode as graphically illustrated in Figure 5.25.

Table 5.2 shows an unusual phenomenon of 0% MSE at m=128 and m=256 (widths of 'chirp' pulse used), when SNR=10 (decimal units) and above. This is because at these instants 100% of the signal is recovered. However, when m=512, this model breaks down as a MSE of 20% is achieved. This shows that the optimum results are obtained at m=128 and m=256 i.e. MSE is dependant on width of chirp pulse used.

By setting a threshold value for reconstruction, an error-free signal can be recovered, the process of which has been illustrated in Figure 5.22 confirming that the method can be successfully used for commercial applications. Therefore, the principle of the technique has been incorporated in two-dimensional watermarking described later in this chapter.

5.7 Cross Entropy (cent)

Let us consider h_i to be an array in an image to be watermarked called as template or an original image and h_i to be a similar array in its counterfeit copy or a photocopy of the original. Then we can say E (cent) which is a value between 0 and 1 is given by:

$$\underset{0<E<1}{E(cent)} = \frac{\sum_i h_i \log(\frac{h_i}{\hat{h}_i})}{\sum_i h_i \log(h_i \hat{h}_i)}$$

$$= \frac{|\sum_i h_i \log h_i - \sum_i h_i \log \hat{h}_i|}{|\sum_i h_i \log h_i + \sum_i h_i \log \hat{h}_i|}$$

The cent can be described as a discrimination information function; it becomes particularly important in the case when we are able to guess an initial distribution of a function before calculating a more precise one.

5.7.1 Resolution dependant Microbar Detection using flat bed scanner Epson 1240U

A photocopy of a Microbar encrypted image shown in Figure 5.26 (the area scanned highlighted with blue border) has been scanned with varying resolutions from 100dpi to 400dpi. An interesting observation was that the Microbar could not be detected at resolutions lower than 300dpi. The average value of the cent for Microbar detection was found to be 0.0003197 (at 300dpi).

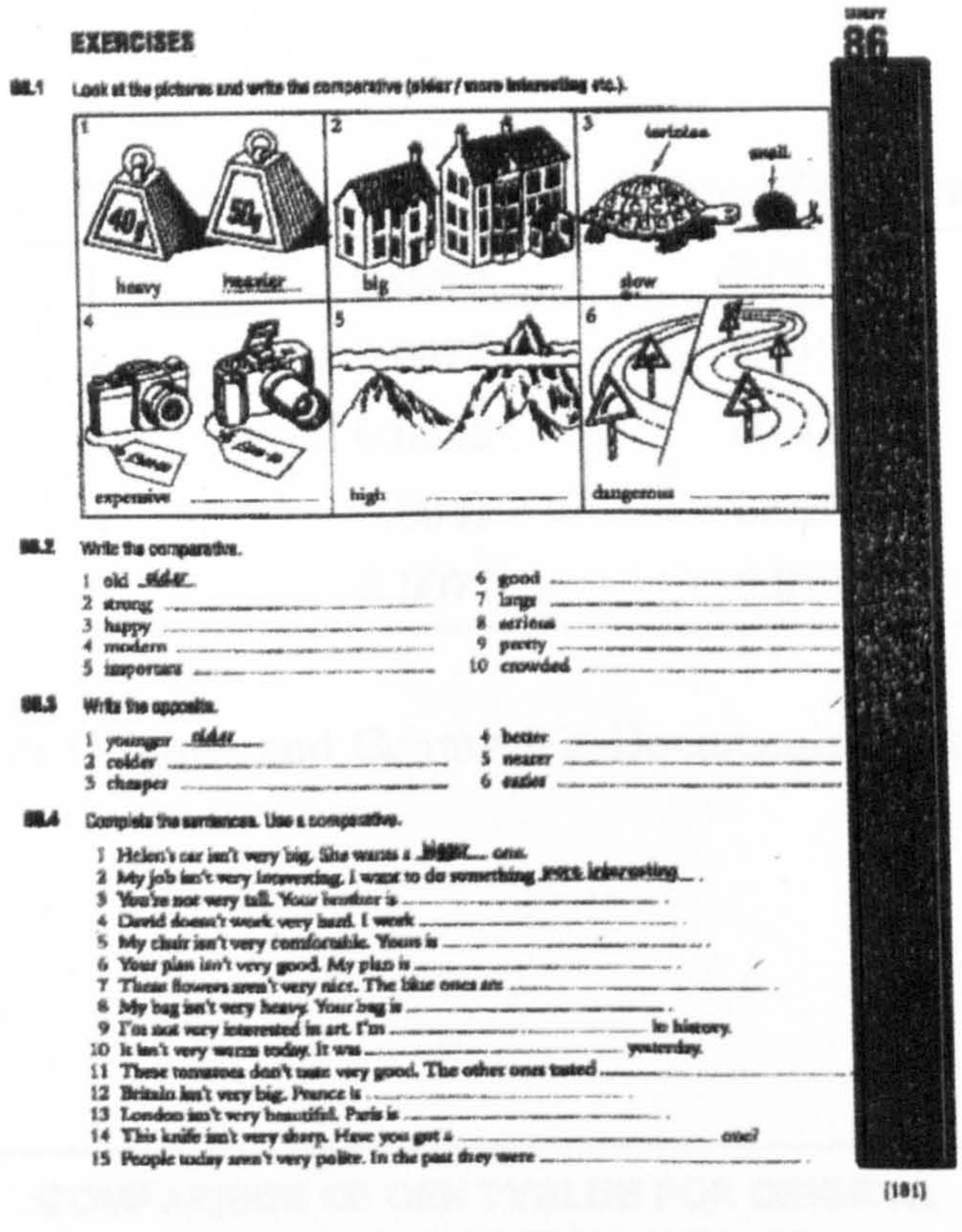


Figure 5.26: Photocopy of a Microbar Encrypted Image

An original and its photocopy (counterfeit) were scanned a number of times (average of 5 is taken in the results shown) and the cent value (cross entropy) displayed on the pen was recorded. The pen has to be held at an angle of 75 to 90 degrees with the rollers lightly touching the page. On plotting the relationship between the number of scans and the cent value in each case, we find that the cent value is higher for the counterfeit document. Therefore, based on the cent value, counterfeit documents can be instantly differentiated from the original ones, making the process very simple and straightforward.

No. of Scans	Original Cent	Counterfeit Cent
1	0.0085	0.0119
2	0.0077	0.0126
3	0.0068	0.0171
4	0.0041	0.0200
5	0.0077	0.0110

Table 5.3: Original and Counterfeit Documents Cent Values

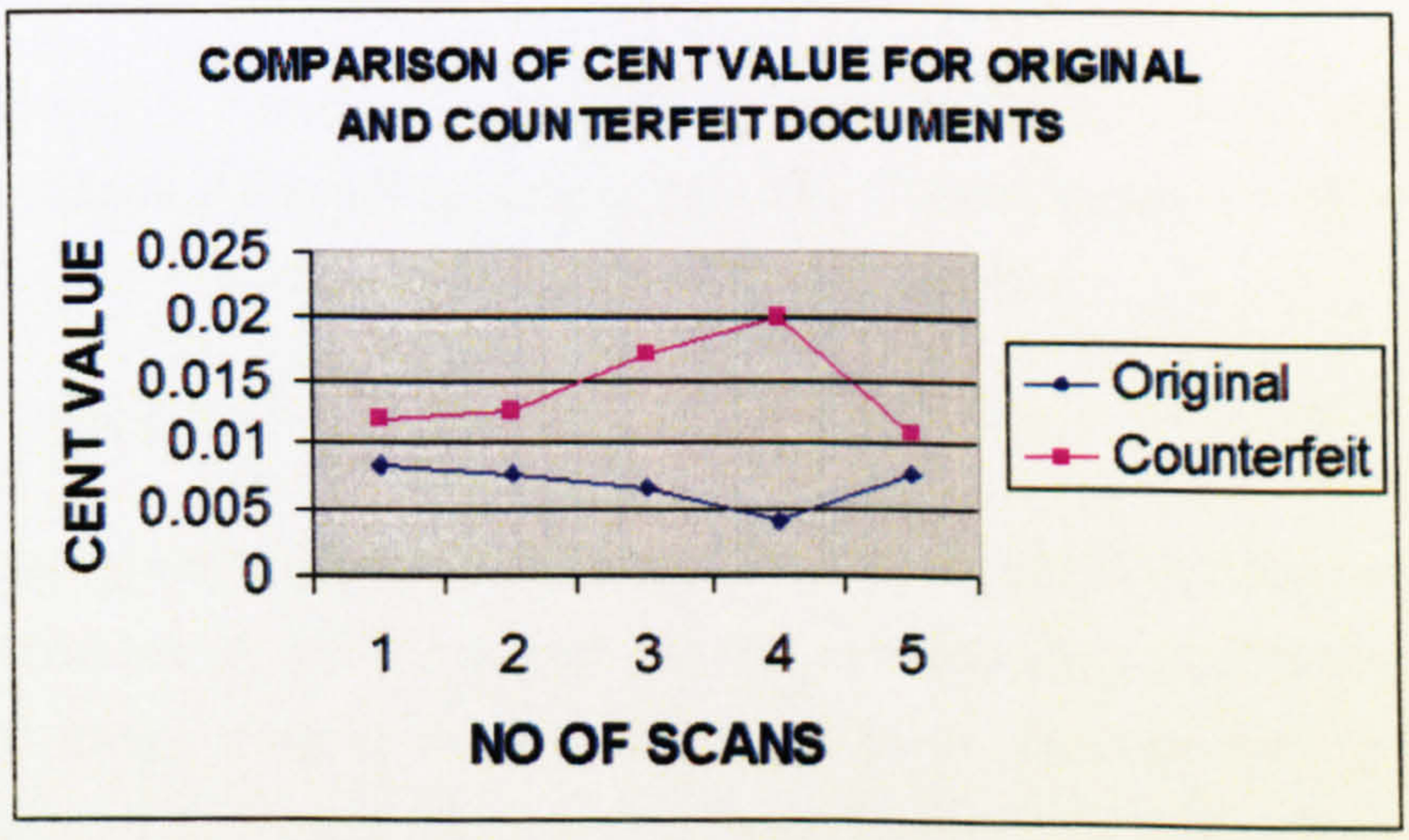


Figure 5.27: Result: Comparison of Cent Value for Original and Counterfeit Documents

5.8 Two-Dimensional Watermarking: Full Digital Watermarking

5.8.1 The Principle

- Same principle, but two-dimensional instead of one-dimensional
- Uses Fresnel zone patterns for coding instead of one-dimensional 'chirp'
- Fresnel coded image used as watermark
- Reconstruction obtained by correlating with Fresnel zone code (password protected)

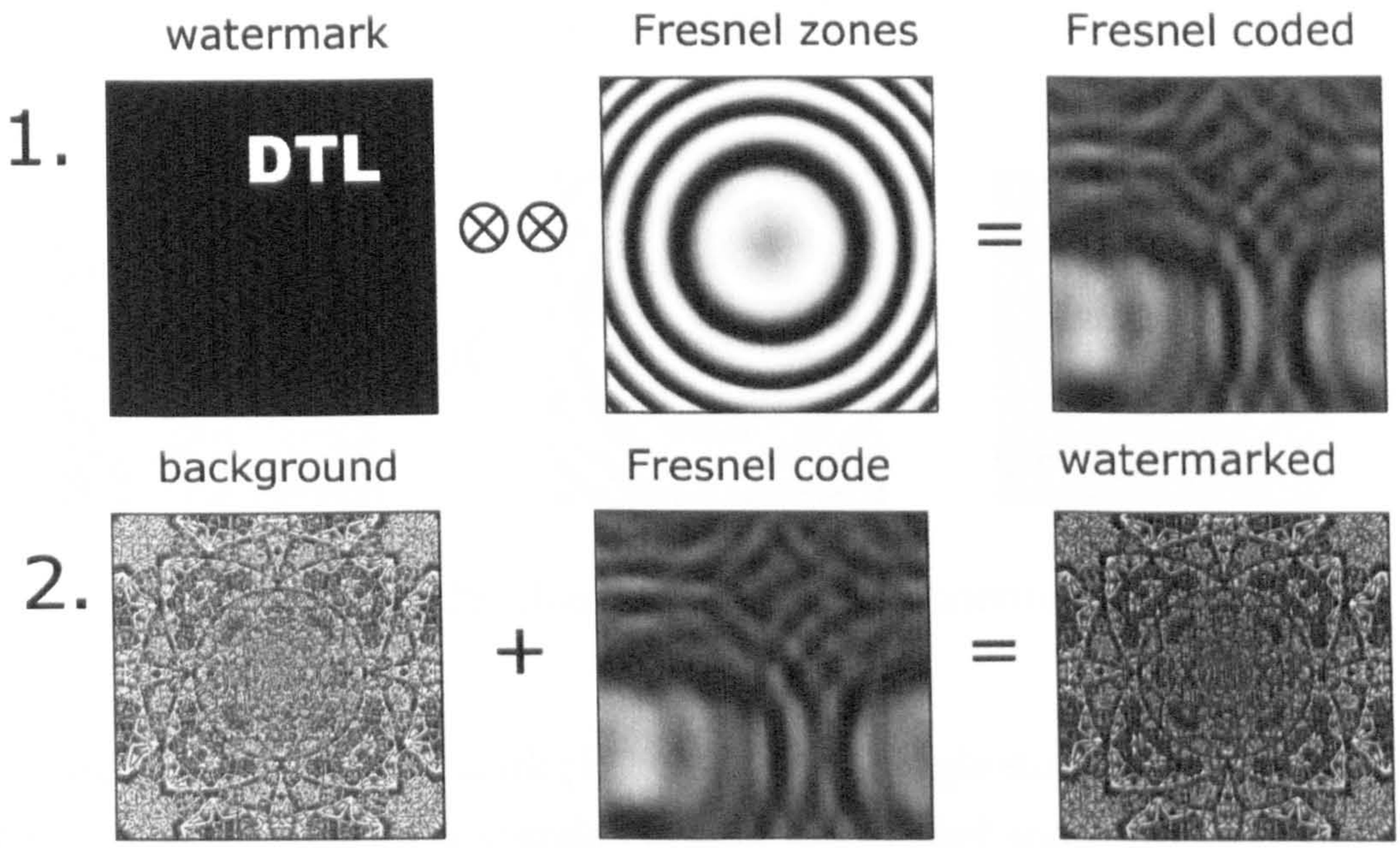


Figure 5.28: Generation of the 2-D Watermark

Referring to Figure 5.28, the DTL watermark is convolved with Fresnel zones to obtain a Fresnel coded image. The Fresnel zones are generated by computing the following equation for two-dimensions

$$\varphi_{(x,y)} = \int\int dx'dy' \exp\left[ik\left((x-x')^2 + (y-y')^2\right)/2R\right]$$

where x and y are original image arrays, x' and y' are shifted image arrays, k is the wavenumber ($k = 2\pi/\lambda$ where λ is the wavelength) and R is the distance of shifted image from aperture. An image to be watermarked (referred to as background in Figure 5.28) is then superimposed onto the Fresnel coded image resulting in required watermarked image.

5.8.2 Recovering the Watermark

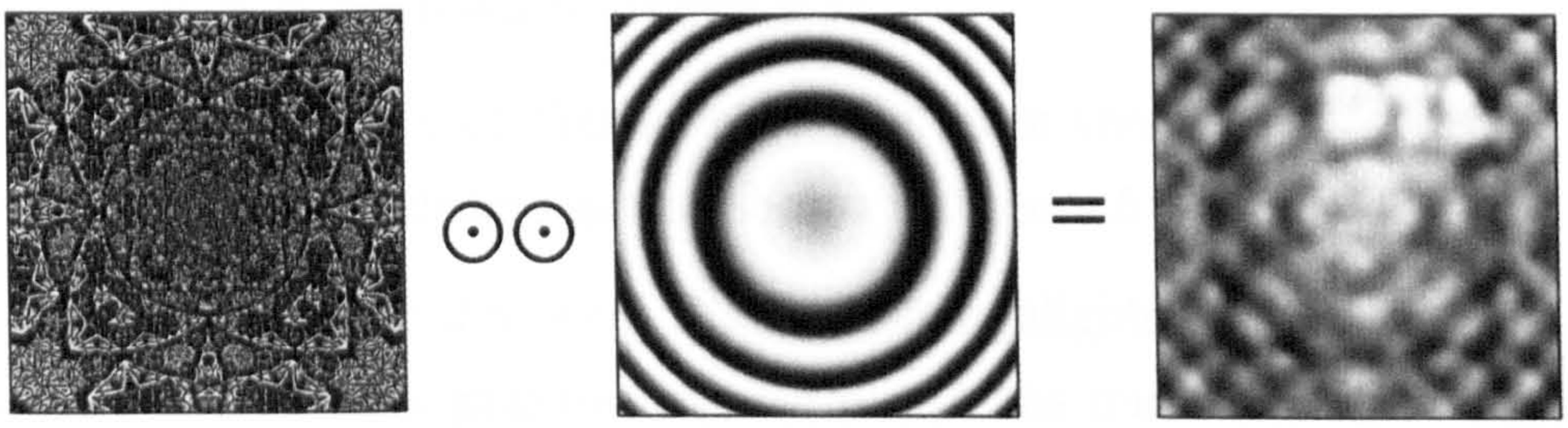


Figure 5.29: Recovering the 2-D Watermark

Recovering the watermark (Figure 5.29) is a single stage process, wherein the watermarked image is correlated with the Fresnel zones resulting in the watermark being detected.

5.8.3 Two-dimensional Fresnel-based Watermarking Analysis

Watermark using Fresnel rings fails when the number of rings are more than the size of the watermark. For example, if $n = 185$ and $m = 200$, where n and m are square array dimensions of watermark and fresnel rings respectively. This is shown in Figure 5.30.

For $SNR < 0.6$ (signal-to-noise ratio): The background image overwhelms the Fresnel coded image, hence attempts to recover the watermark fail as shown in Figure 5.31.

For $SNR > 0.6$: The Fresnel coded image progressively appears to be at the forefront of the background image. However, recovery of the watermark is comparatively good as highlighted in Figure 5.32.

When the number of Fresnel rings are less than the ideal value (in this example the ideal value is $n = 40$), the recovered watermark appears progressively blurred as shown in Figure 5.33.

When the number of Fresnel rings are more than the ideal value, the watermark gradually disappears as shown in Figure 5.34.

For $SNR = 0.6$ and $n = 40$: Figure 5.35 highlights the ideal balance (for this example) of the signal-to-noise ratio and the number of Fresnel rings. The mean square error (MSE) percentage between the recovered watermark and the original watermark is approximately 45.65%. This value though high, is a trade-off between percentage of recovery and the covert nature of the watermark. Table 5.4 shows the error percentages when the number of Fresnel rings, n and signal-to-noise ratio are varied.

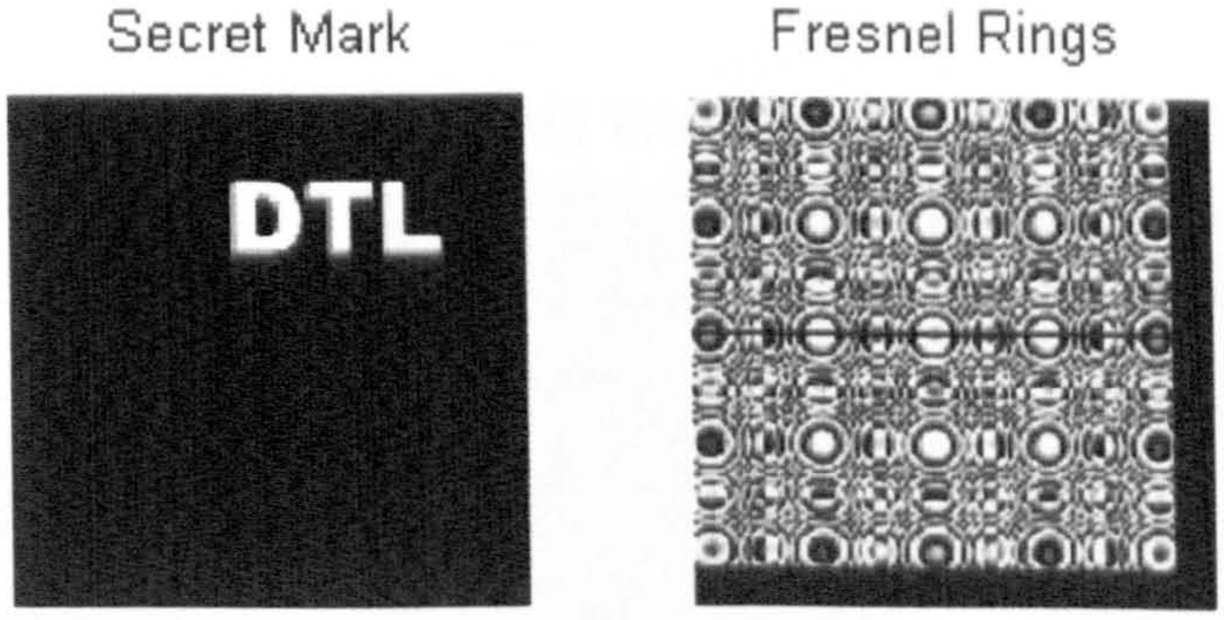


Figure 5.30: Fresnel rings image larger than watermark

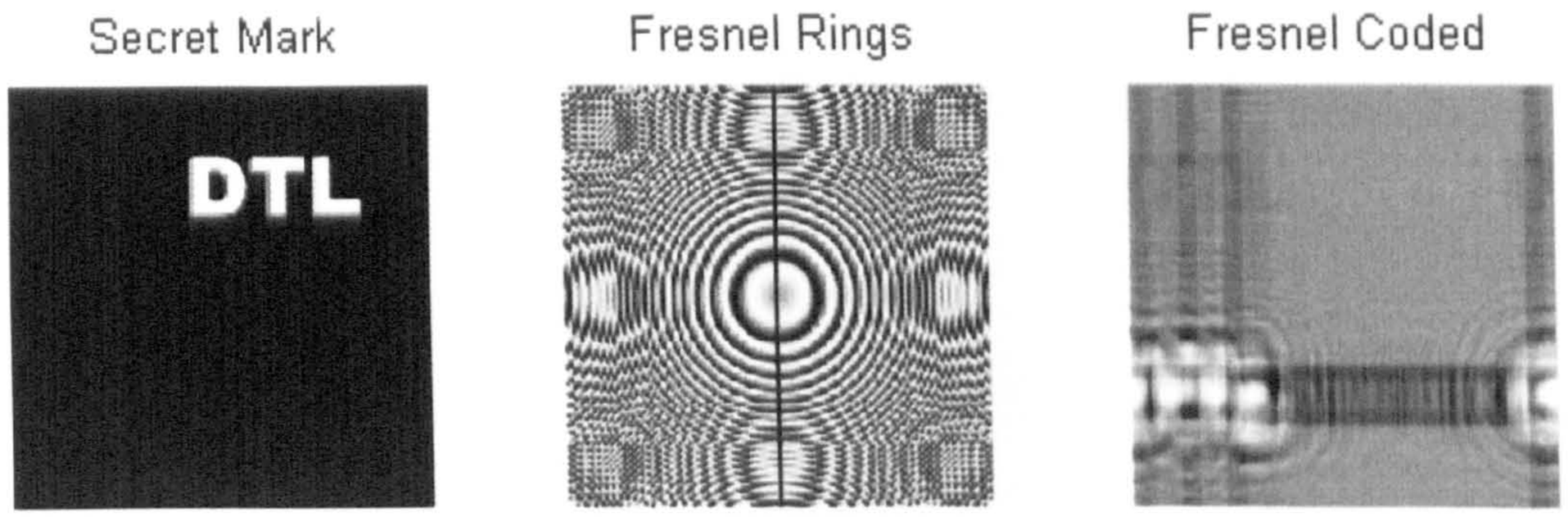


Figure 5.31: For $SNR < 0.6$

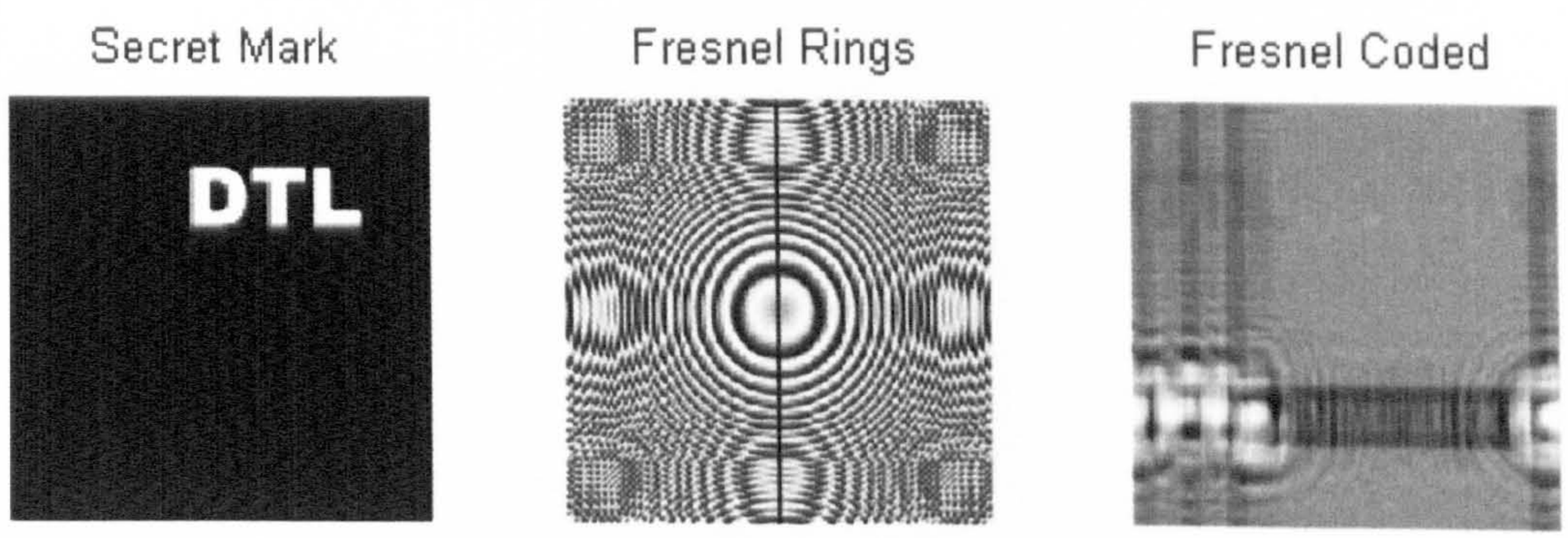


Figure 5.32: For $SNR > 0.6$

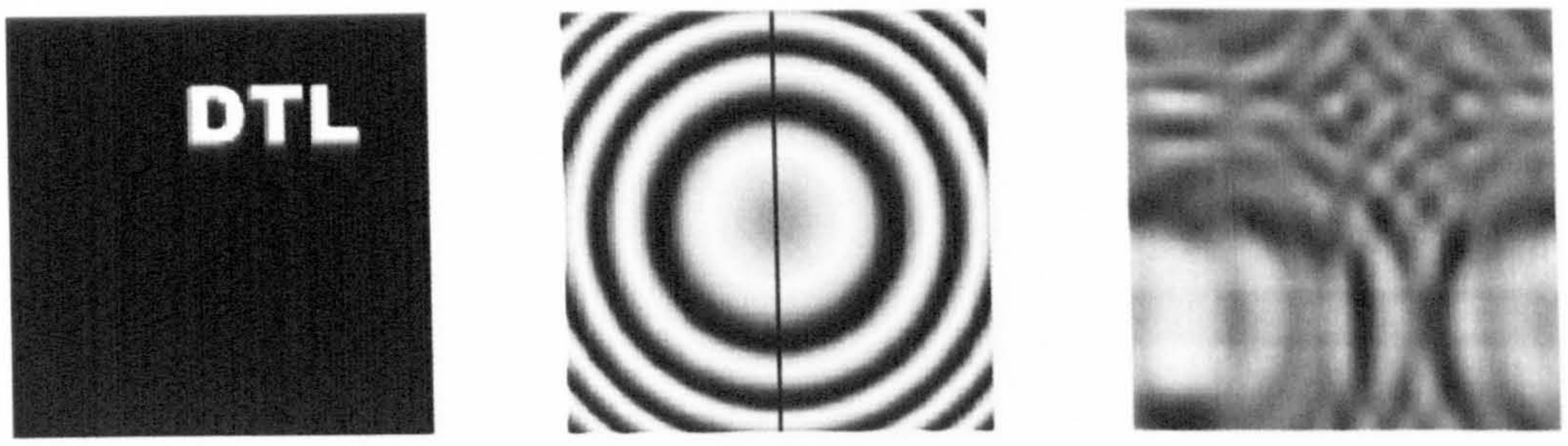


Figure 5.33: When Fresnel rings, $n < 40$

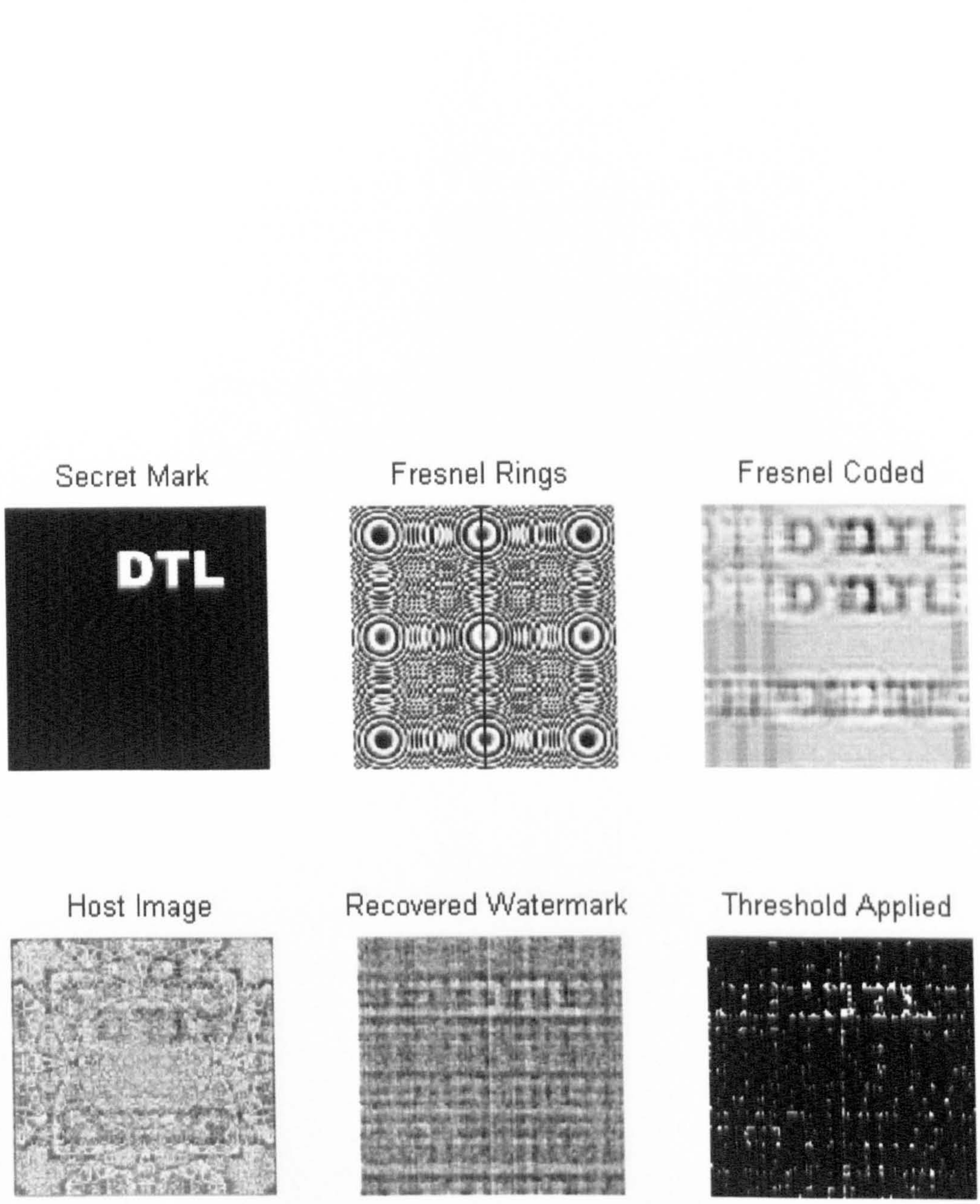


Figure 5.34: When Fresnel rings, $n > 40$

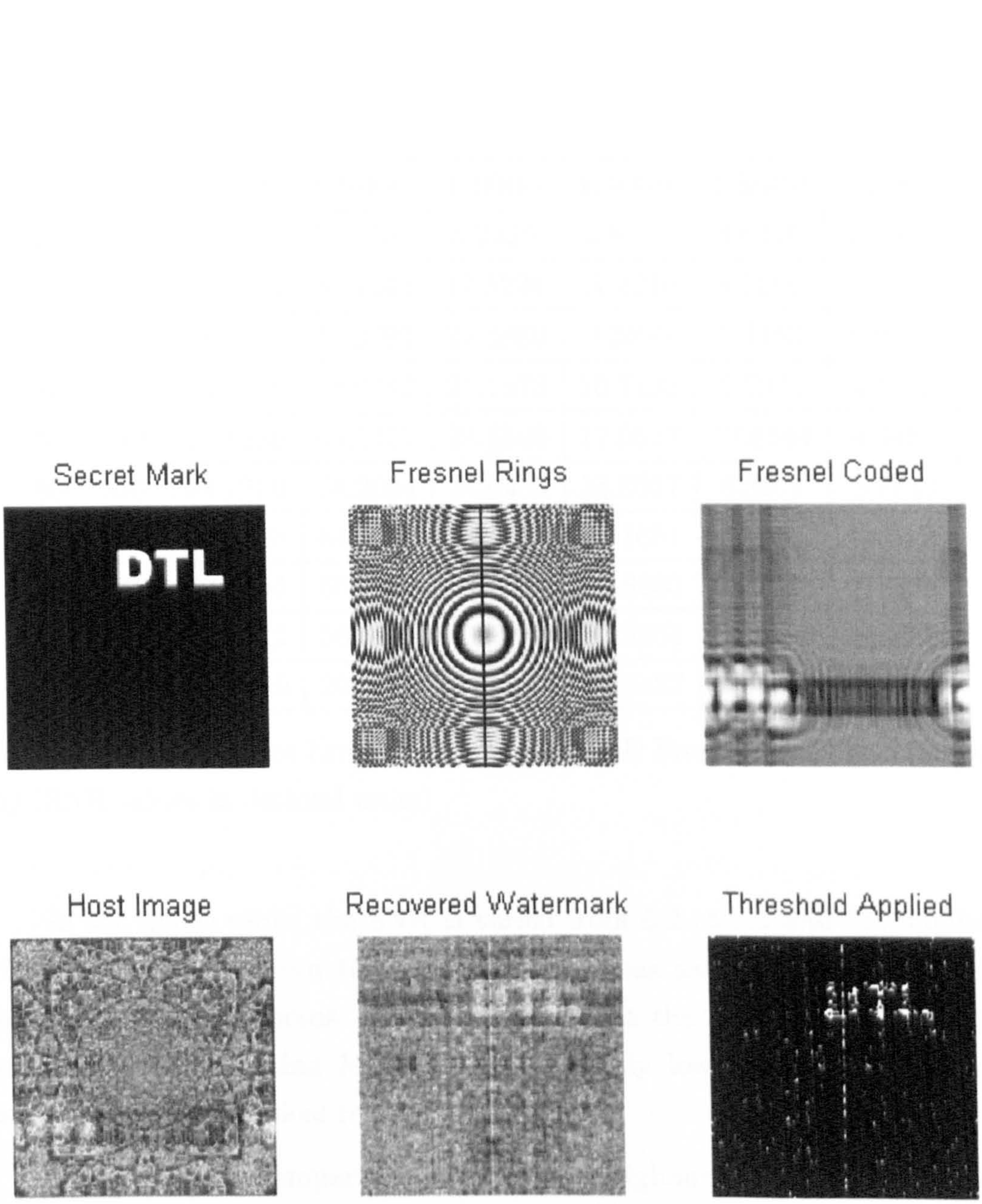


Figure 5.35: $SNR = 0.6$, $n = 40$

n/SNR	**0.2000**	**0.6000**	**1.0000**	**1.4000**	**1.8000**	**2.2000**
10.0000	62.8214	12.3258	5.2228	3.8132	3.6919	3.5532
20.0000	75.1785	36.7246	17.5224	11.4210	9.1586	7.5752
30.0000	84.9242	44.0793	22.8880	9.3490	5.7453	6.0468
40.0000	83.9975	45.6482	21.1878	10.7106	5.7975	4.1496
50.0000	93.1230	64.1557	24.5598	17.9517	12.8544	4.3487
60.0000	86.7759	58.3094	24.0453	18.8017	8.3377	5.7849
70.0000	92.8398	64.3518	29.1173	21.1651	15.4439	12.0450
80.0000	93.3494	66.6484	39.2936	29.8383	22.1633	21.4606
90.0000	81.9512	56.6206	41.7738	35.5054	31.9598	30.3928
100.0000	65.8219	20.9614	14.6163	16.0427	13.4982	11.9768

Table 5.4: Mean Square Error percentages for 2-D Fresnel-based watermarking (SNR values in decimal units)

In the above table, the SNR is varied from 0.2 to 2.2 and the number of Fresnel Rings, n, from 10 to 100. These values have been chosen purely for experimental purposes. It can be seen that the watermark covertness increases with increasing MSE and consequently lower SNR's which is in accordance with practical results.

An exhaustive comparison between the Digimarc and Microbar technologies cannot be made for the purpose of this thesis due to the Digimarc copyright issues.

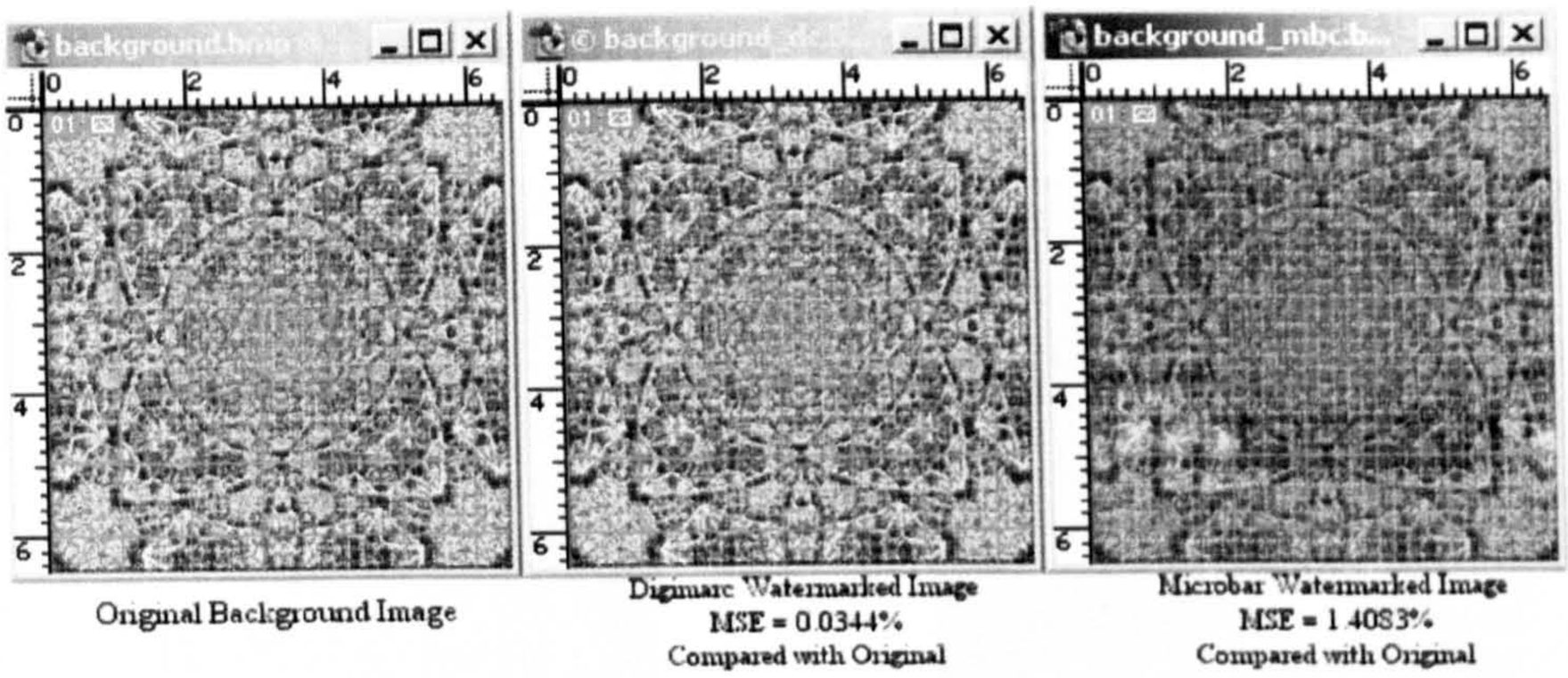


Figure 5.36: Microbar and Digimarc Watermarked Image Comparison

With reference to Figure 5.36, a comparison between an original image and a watermarked image with respect to the mean square error has been made using both technologies. It has been observed that there is less degradation of the watermarked image compared to the original when watermarking using the Digimarc technology. However, the difference in error using the Microbar technology (larger compared to Digimarc watermark) can be attributed to other factors, for example for Microbar technology, the watermarked image has been created in MATLAB (with un-optimised code) and has been converted twice from a greyscale to indexed to RGB (Red-Green-Blue) with 24-bit depth image, while the Digimarc watermarked image is created within Adobe Photoshop (an optimised commercial application). Furthermore, the Microbar watermark used in this example is the DTL text image while within the Digimarc watermarked image is embedded the Digimarc ID and copyright year. Hence, a direct comparison between the two techniques cannot be made.

5.8.4 Example: Logo-type Watermark (with PIN)

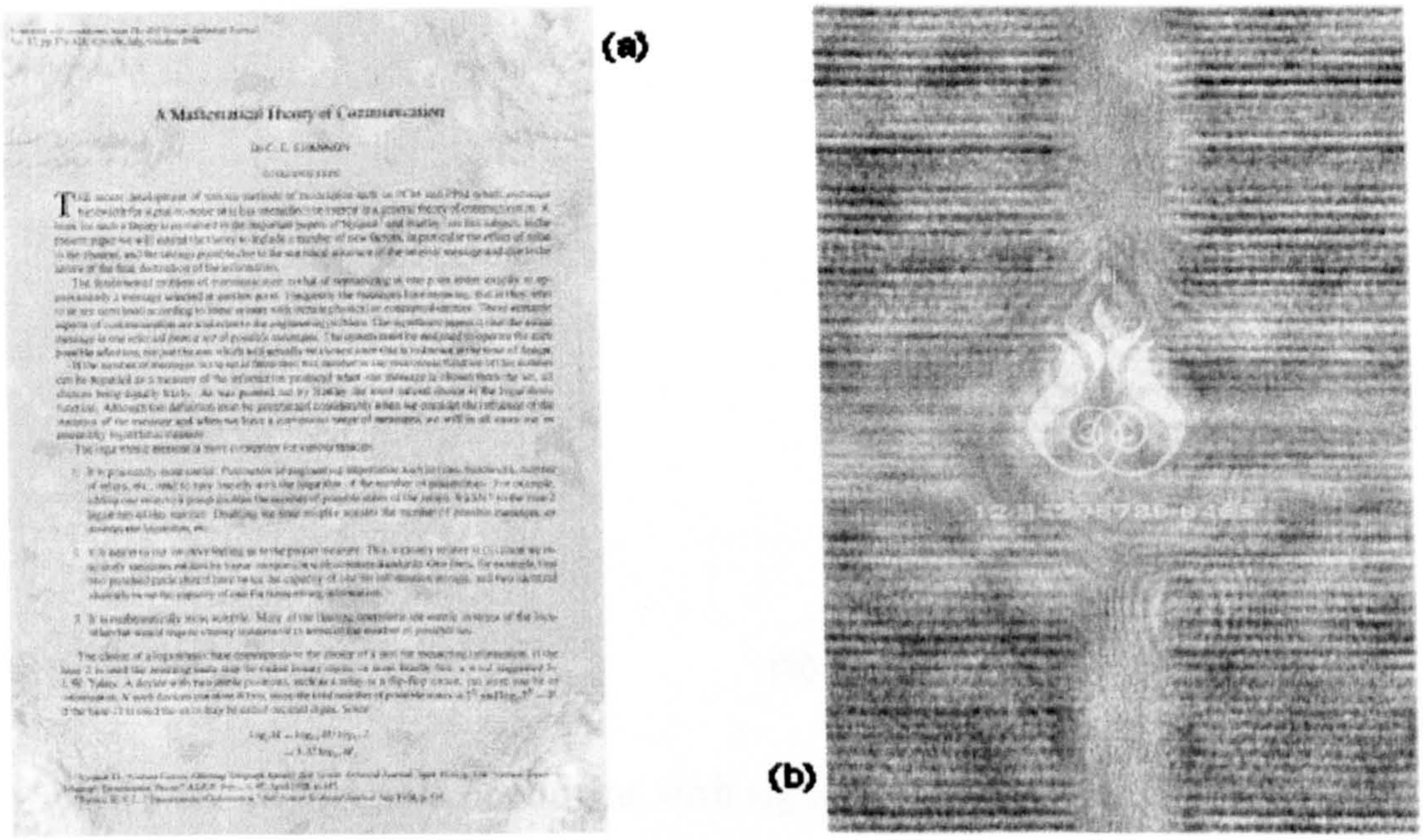


Figure 5.37: (a)Document with invisible watermark (b)Recovered Watermark

In the above example, Figure 5.37(a) shows a document with the 2-D watermark embeded within it and is invisible due to its covert nature. Using the same principle explained earlier, the watermark has been recovered in Figure 5.37(b). The principle works with signature types watermarks as well as shown in Figure 5.38(a) and 5.38(b).

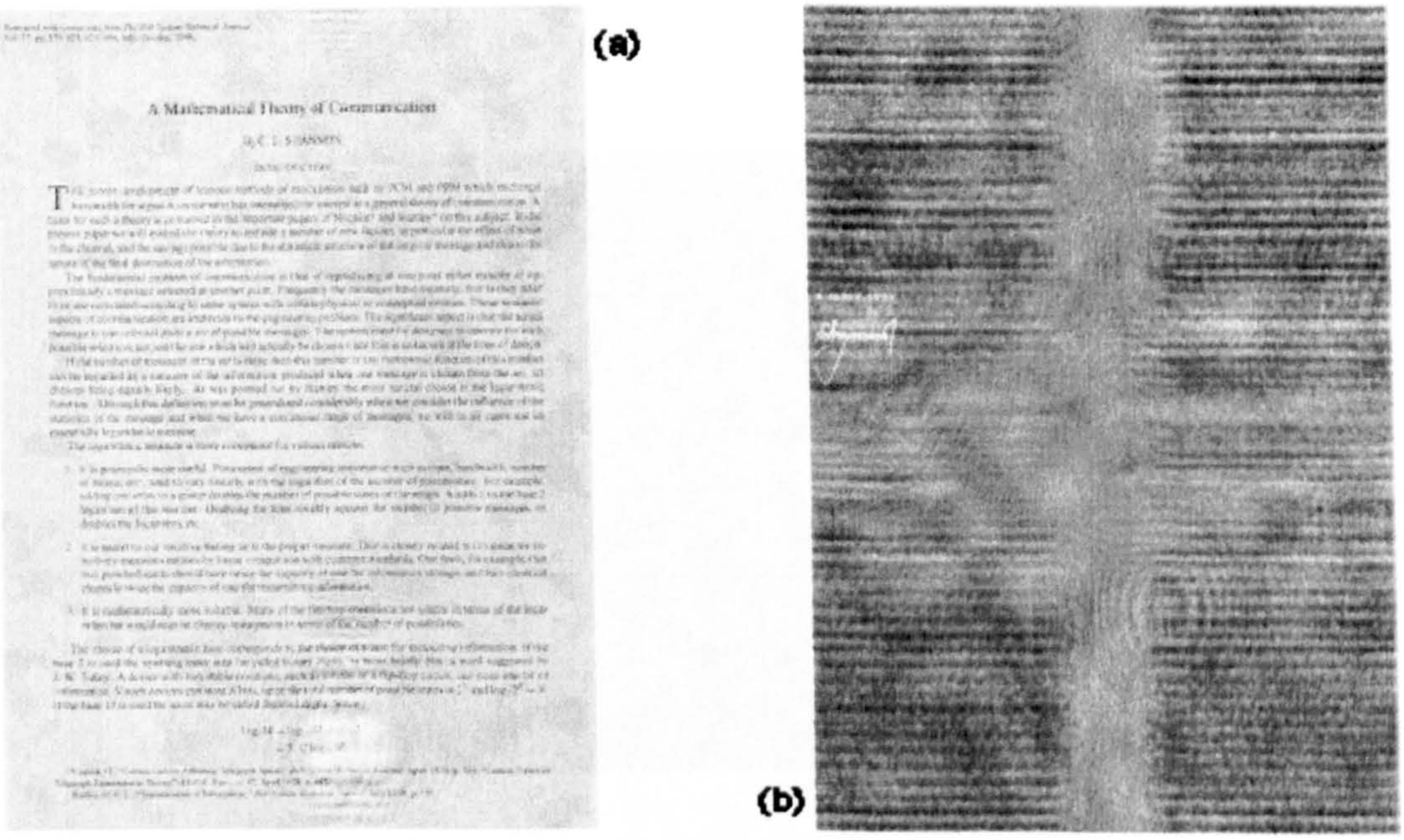


Figure 5.38: (a)Original document with signature type watermark embedded (b)Signature watermark recovered

5.9 Fractals and Camouflage

This is a conceptual example of how we hide 'Microbar' in a textured image. Figure 5.39 is a fractal field made up of grey scale values from 1-255. It has the appearance of a rouch surface, like a concrete slab. The 'roughness' is termed as the fractal dimension which is a measure of the 'roughness'. It can be seen that the top right hand corner of the image appears different to the rest of the image. This is because data has been taken out and replaced with completely different information like putting the Microbar signature in an image to be protected. It becomes visible because its fractal dimension is different to that of the rest of the image. If the fractal dimension is tuned such that it matches the rest of the image, it can disappear as shown in the

Figure 5.40.

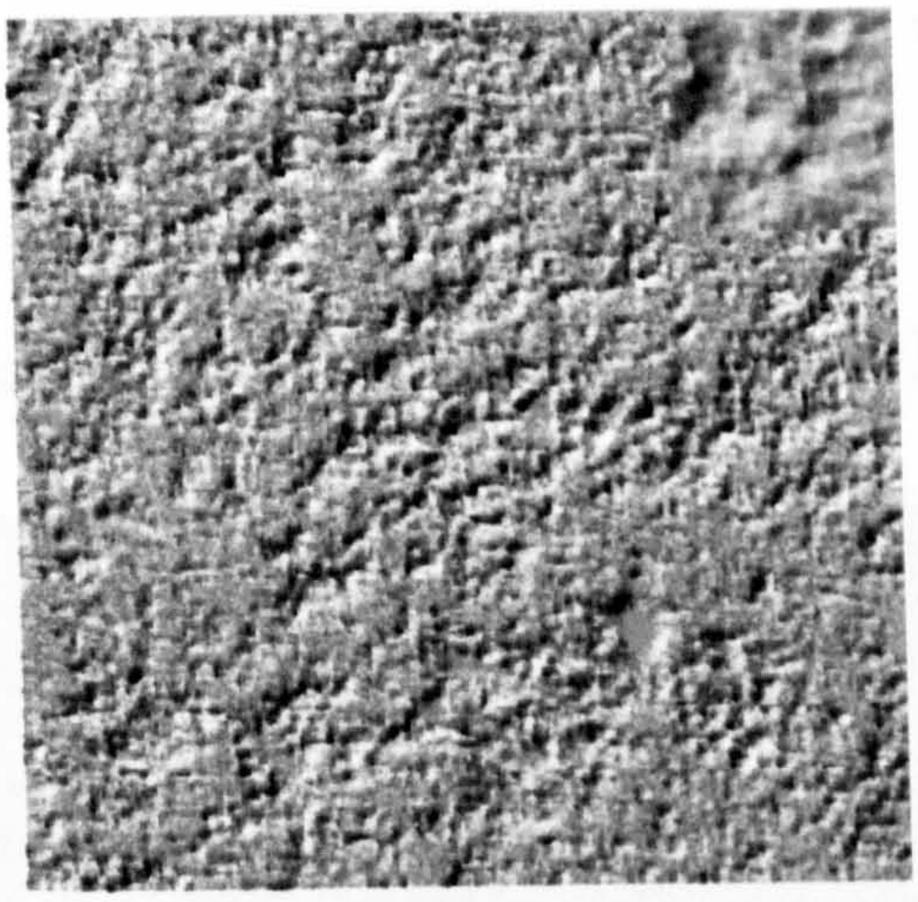


Figure 5.39: Example: Microbar hidden in Textured Image

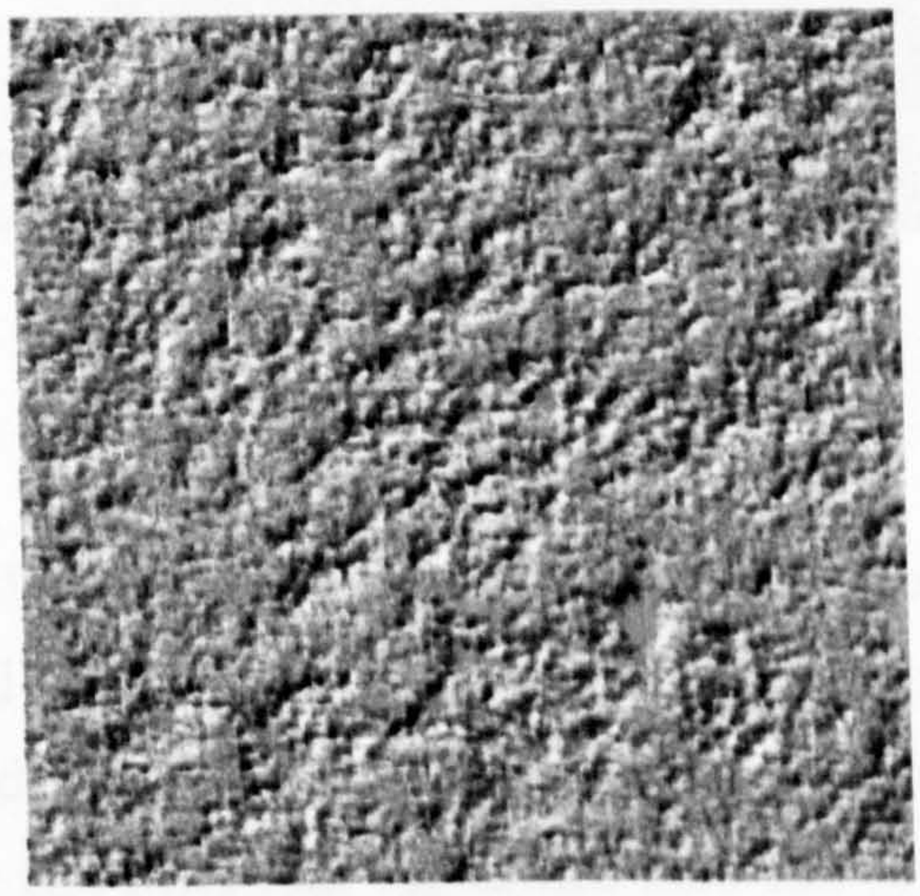


Figure 5.40: Microbar camouflaged with the rest of the Image

Figure 5.40 shows that by tuning the fractal dimension to match the rest of the image, the new information is difficult to discern from its surroundings.

It is still there, still different, but our eye-brain interface is deceived by the 'natural camouflage' of fractals. A similar technique is used in the wild, where a leopard's spots and a tiger's stripes blend into the dappled sunlight and shadows created by the natural vegetation of their habitats.

5.9.1 Example: Picturesque Watermark

Figure 5.41: Fractals in Nature

The Figure 5.41 shows how prevalent fractals are in nature. The left figure is an 'original' of a mill stream flowing through woods. The centre figure shows a 'cloud based' Microbar signature. The right figure shows the result of embedding the Microbar signature into the picture at 50% opacity. This high opacity value means that the Microbar has been introduced into the 'original' picture at a level of 50% of the original picture. Typically, we embed signatures (watermarks) at anywhere between 10% - 30% depending on the nature of the image being protected. This shows that the Microbar watermark is very robust and the technique is successful even at high opacity.

The difficulty of seeing the very high level of Microbar (especially within the trees element of the picture) shows how strong a 'fractal' the trees are in the first place. This re-enforces the view that:

- We are, unknowingly, already surrounded by 'fractals'.
- 'Natural' images provide a perfect backdrop for Microbar.

5.10 Application to Document

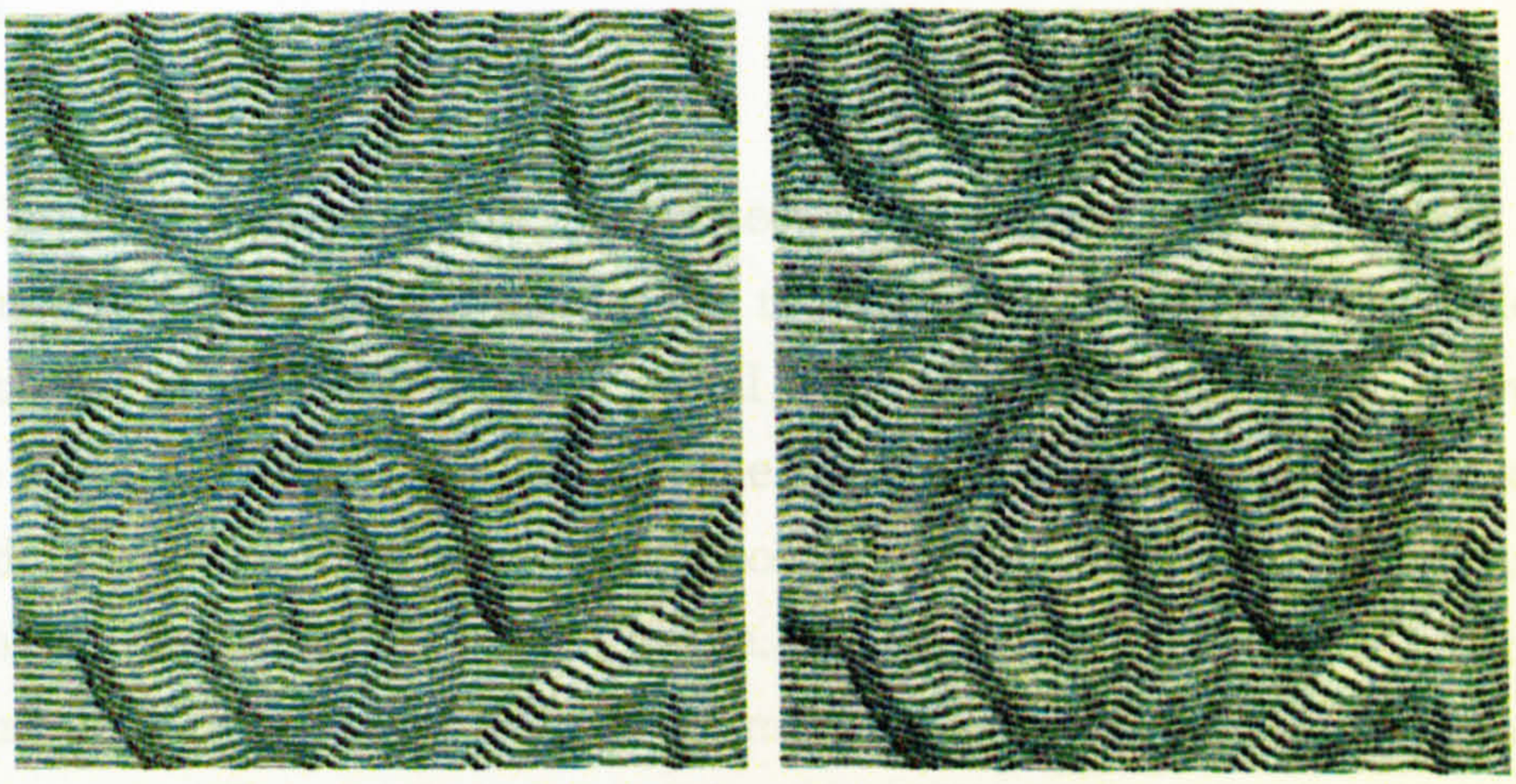


Figure 5.42: Left:Non-encrypted Image; Right:Encrypted Image

In practice, the presence of the Microbar signature within a binary image is not discernable to the human eye. The Figure 5.42 above has been magnified to show the effect but even then we can see that the change is minimal. It is only when we magnify the image to extremely high levels that we can see the Microbar signature, though even here the effect is somewhat exaggerated as shown Figure 5.43.

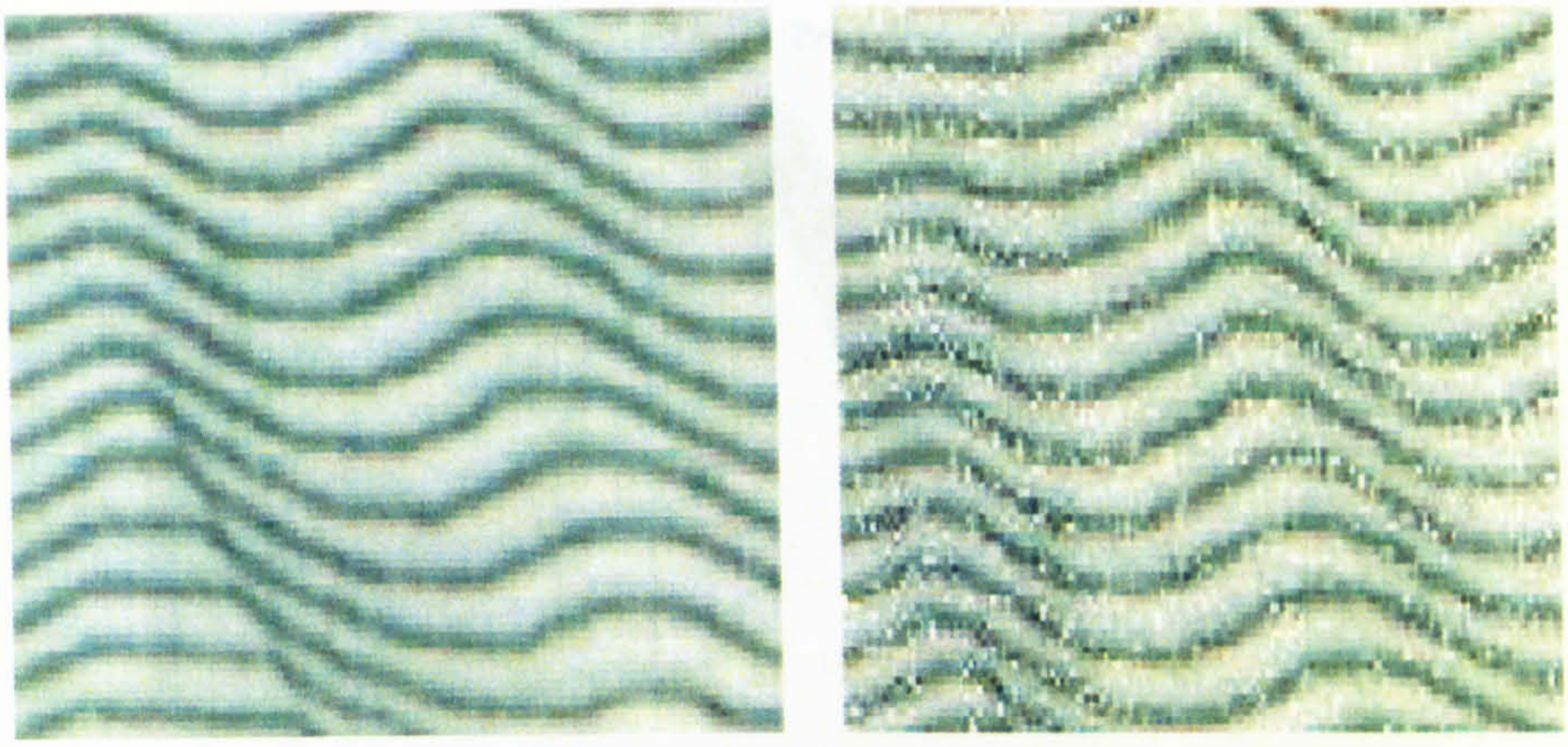


Figure 5.43: Left:Zoomed Non-encrypted Image; Right:Zoomed Encrypted Image

The Microbar works for the protection of Binary Images as well. With reference to Figure 5.44, imagine the rectangle on the right is a whole document onto which a Microbar fractal signature has been applied over the whole surface. Now, imagine one line from a typical security printed 'fine line' detail image. This line then becomes 'peppered' with the white dots making up the Microbar fractal image. When the image is printed, a certain amount of 'blur' occurs (no printing machine will completely replicate a digital image). When such a printed document is then scanned, more 'blur' is created (no scanner is a 'perfect machine; irrespective of the resolution). As a result we have two degradations of the image when a document is read and we have created our own grey-scale image in the process. When a counterfeiter attempts to copy and print a genuine version, further degradation and associated 'blur' will occur. We essentially compare the levels of grey-scale or 'blur' of the suspect document with the scanned image of the original.

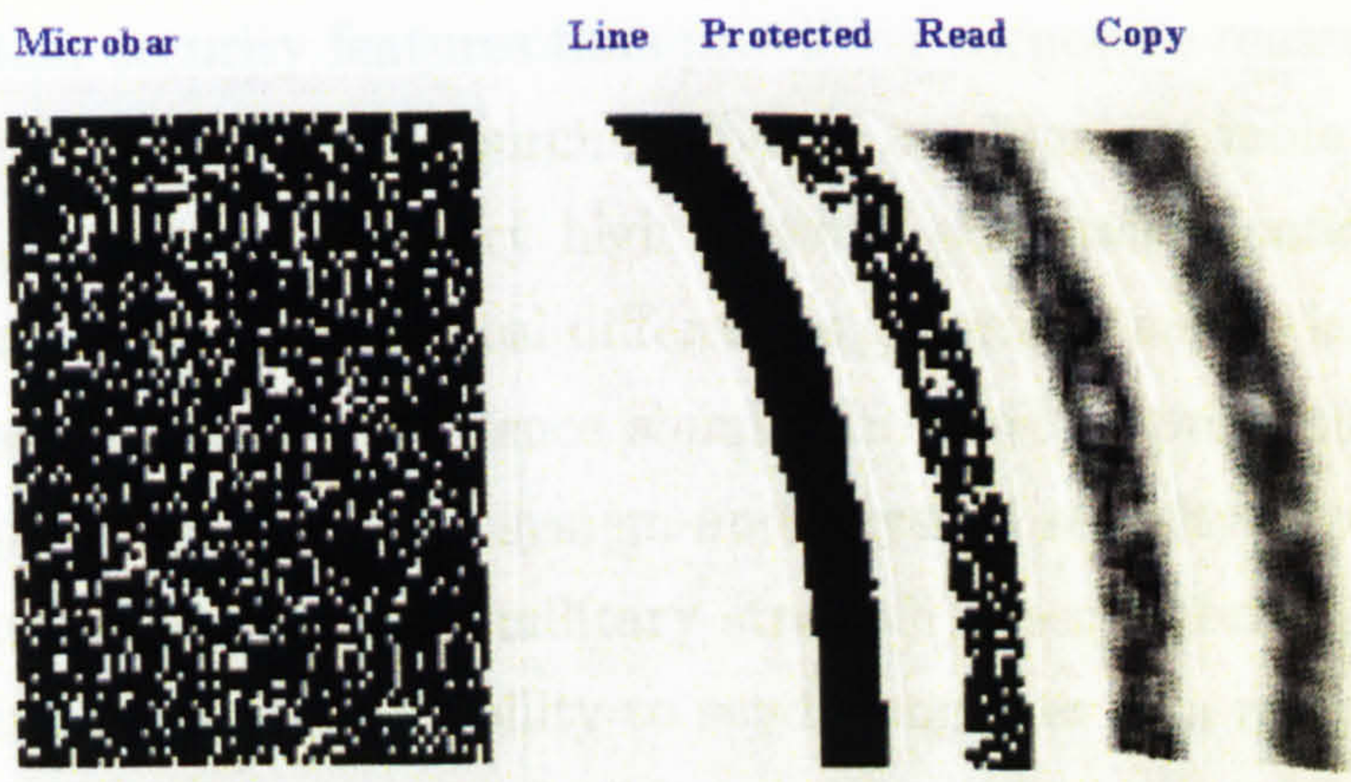


Figure 5.44: Use of Microbar for Binary Image

5.11 Microbar Print: Key Benefits

Figure 5.45: Use of the Pen reader

Thus, it can be seen that the Microbar is very easy to implement being a 'print feature', users can continue to print in exactly the same way as before. Its covert nature allows incorporation into existing designs negating the need to create additional space on a document for authentication as required by holograms or other overt security features. It can however work

alongside other security features thus providing corporate reassurance where policies do not allow single sourcing. Being machine readable and statistically based, it can provide very high levels of statistical confidence around accept/reject criteria. Statistical differences generated are so large that they could be used as forensic evidence should the need to prosecute arise. The security is paramount as both system and physical security is provided. The system security refers to the (military strength) encryption of the signature on the document as well as ability to send templates to a reader population world-wide. The technique is flexible as the scanning can be done using a conventional flat-bed scanner good for high-value documents and other authentication which would be naturally 'office-based' or a discreet portable pen as described earlier useful for those applications where there is a need to authenticate labels, documents, ID cards etc. 'in the field' or discreetly in store. Easy scanning (whole or part of the document in any direction) is a result of fractal design. As mentioned earlier, it can be used for authentication, batch tracking and product personlisation. As it requires no special inks, taggants or substrates, a print file is encrypted and printing continues as before. This makes implementation, total costs of ownership much more attractive. Being digitally based, Microbar is suitable for all printing processes using digital print files. It provides a complete solution for optical media in that it works with silk-screen printing for CDs, overt Microbar and the enabled mouse can be used for software unlocking. Unlike overt features, like holograms, which rely on human knowledge and interpretation, Microbar is a machine based authentication system. This removes all the potential problems which the highly subjective nature of human intervention brings. Microbar is objective, giving straightforward pass/fail messages and/or track and trace information.

Whether the flat bed scanner or the pen reader is used, the system works by first loading into the software a 'template' of the expected values generated

by a 'genuine'. Upon scanning, the software (MATLAB code) compares the actual values generated by a 'candidate document' with those of the genuine held within the template. The distribution of these 'templates' can be via e-mail and can be secured through the use of any method.

Microbar is also capable of being read in 'real-time'. This has implications for high speed reading of currency and cheques going through counting and sorting machinery where numbers of items read are measured in feet per second. The following section diagramatically explains the application of Microbar to labels and documents.

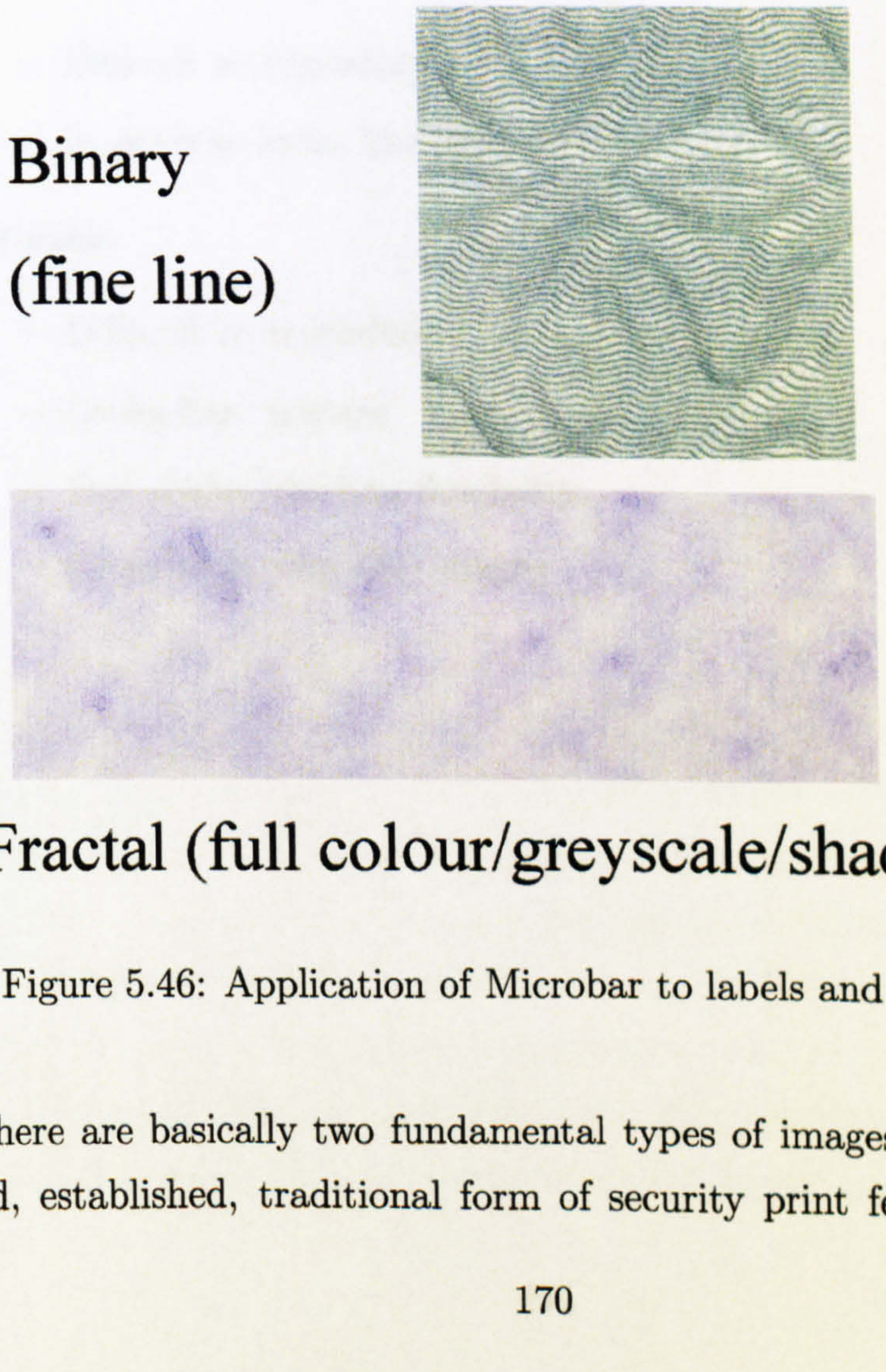


Figure 5.46: Application of Microbar to labels and documents

There are basically two fundamental types of images; binary, which is the old, established, traditional form of security print featuring 'fine line'

details and 'fractal' which uses full colour, shading and grey scale. Traditional fine line images are called 'Binary' because they have either a single colour or are blank. Fractal images by contrast, use shading and a spectrum of colours. Microbar can work with both binary and fractal images, in either a retrospective way, i.e. incorporating into existing designs or can be used during the design stage of new documents, labels etc. The preference is to use Microbar within fractal images, ideally with high degree of texture or complex detail which allows to hide the Microbar signature more completely.

- Fine Line print
  - Difficult to reproduce
  - In practise looks like 'texture'
- Fractals
  - Difficult to reproduce
  - Looks like 'texture'
  - Full colour/shading flexibility
  - Links with 'real life' images

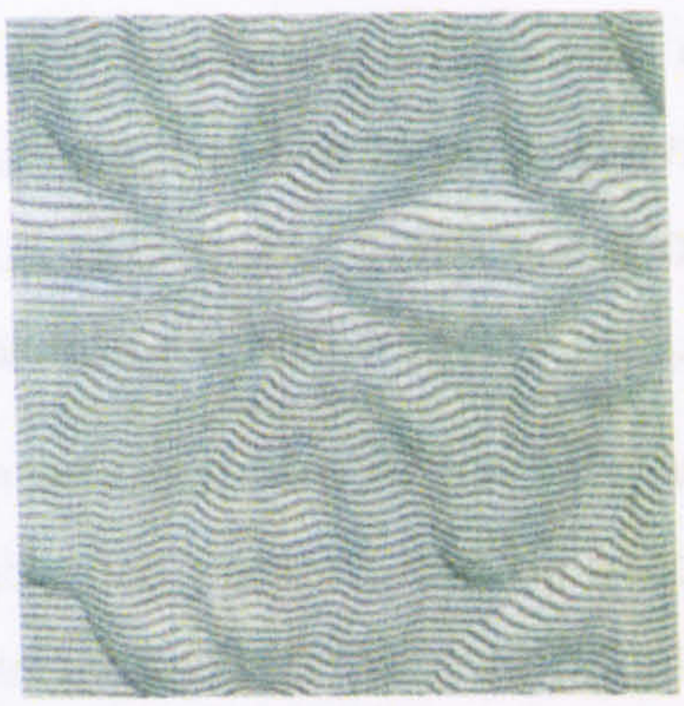


Figure 5.47: Fine Line Print

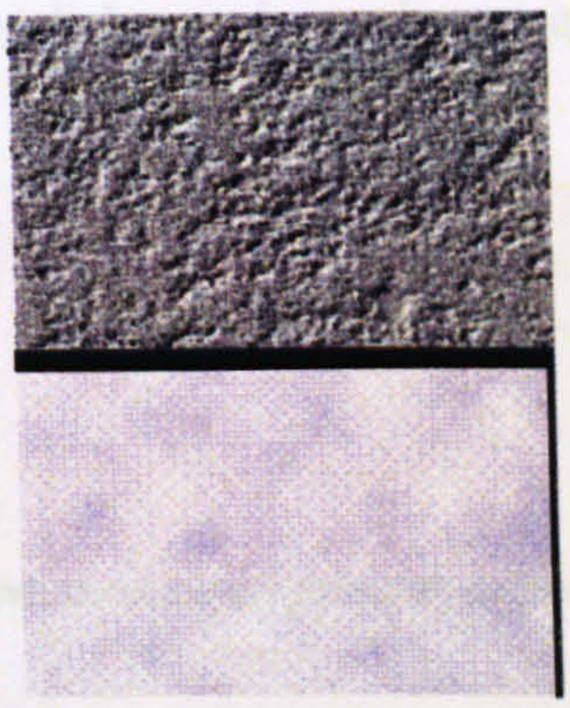


Figure 5.48: Fractals

There is a gradual realisation that the traditional 'fine line print' approach may have had its day. It was originally conceived due to its difficulty to reproduce with the technology available at the time. Interestingly, in practice, the fine line detail actually has the appearance of 'texture' i.e. it often looks like embossing or raised surfaces. In contrast, fractals through chaotic generation are also difficult to reproduce and they certainly cannot be copied. They also have the capability of looking like texture, however, they unlike fine line print, offer the potential of full colour and shading with the

much wider applications and flexibility which this offers. Fractals also have a closer link with natural images, like clouds, flowing water, rippled water, trees, rock formations, etc. Thus this gives the scope to use the technology for a much wider range of images and applications.

Figure 5.49 shows some examples of fractal patterns which can be computer generated. The embossed pattern can be used as a cheque background, the clouds background can be used in many applications. Other patterns, like 'marble' are a 'natural fractal' and can convey a quality image in documents, labels, etc. The choice of 'fractal textures' depends on the application e.g. cheques, documents etc. As all the textures are robust with the watermark being covert, the choice could depend on the aesthetic requirement.

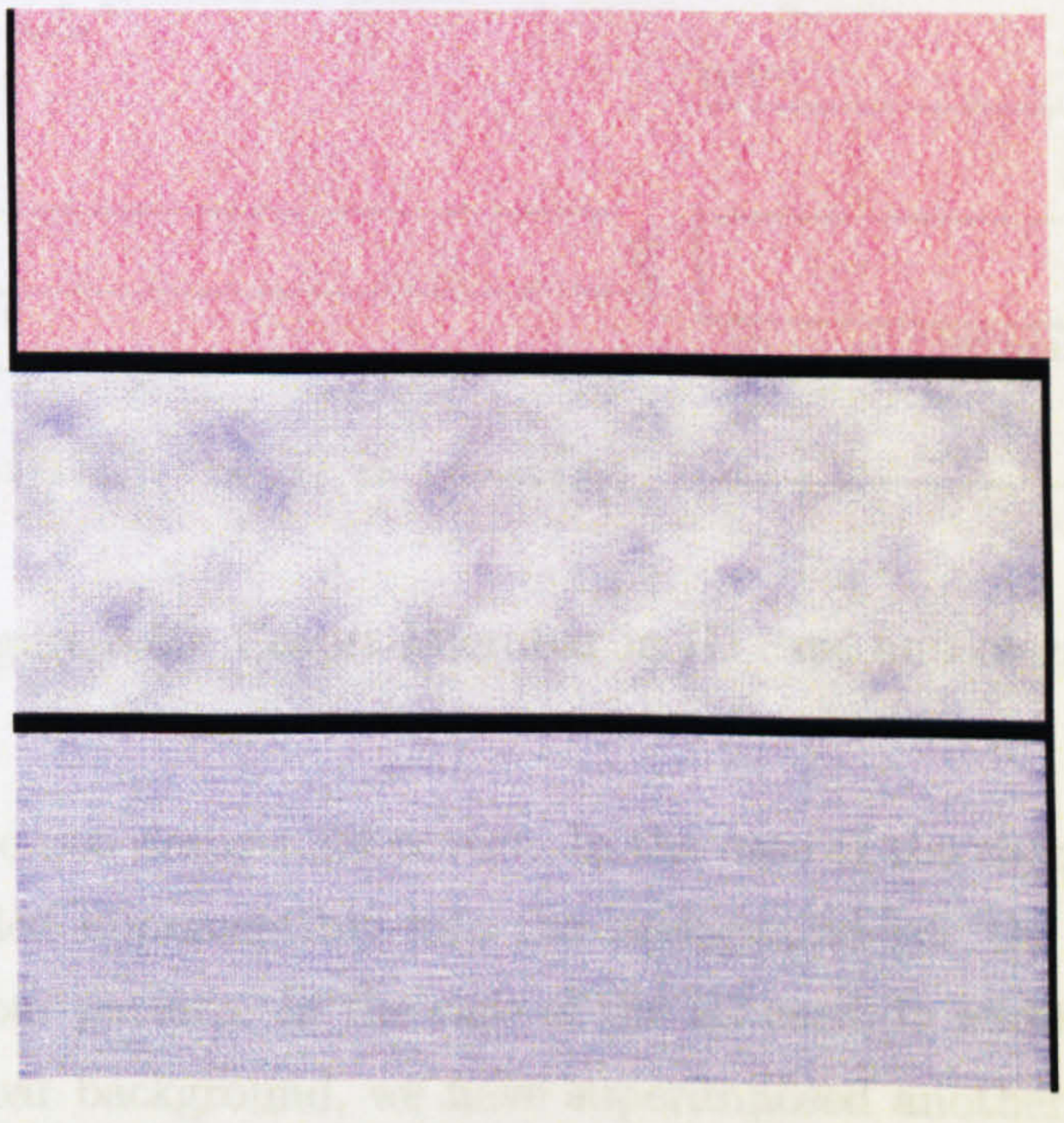


Figure 5.49: Computer Generated Fractal Patterns: cheque background, clouds and flow fields

Below are some practical applications of fractal designs (Figure 5.50). The first one is an identity card and the second is a cheque.

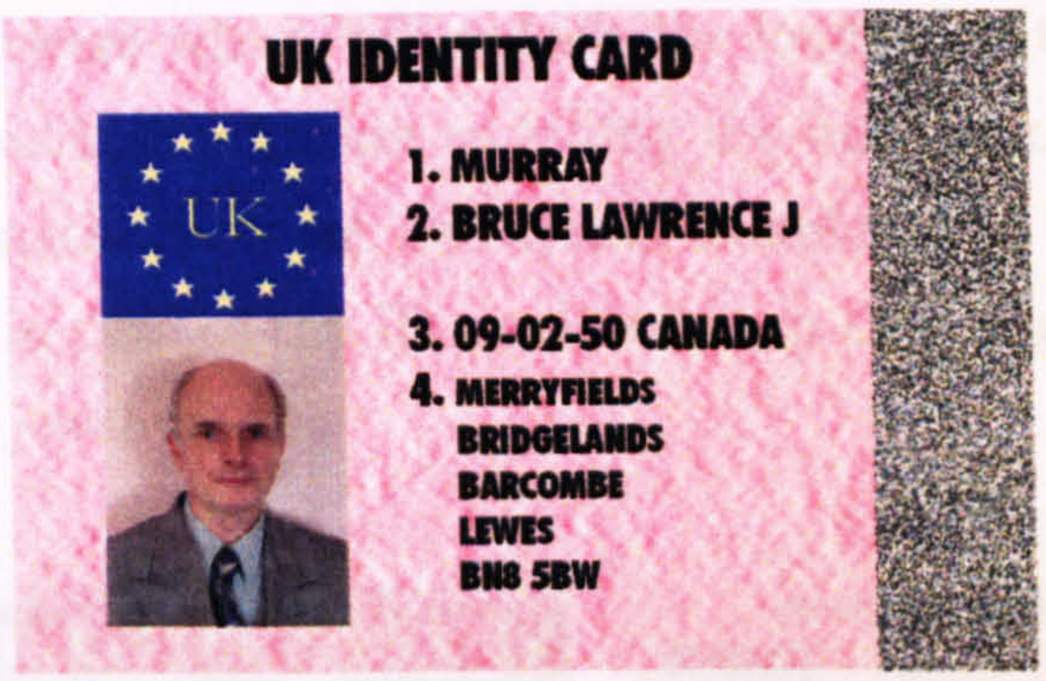


Figure 5.50: Covert Microbar in ID card and cheque

The above two designs are covert. In the case of the cheque, the Microbar is embedded imperceptibly into the background i.e. the background is itself a Microbar pattern. In the case of the ID card, in addition to the embossed Microbar background, we have superimposed another Microbar into the photograph and flag. However we can have them visible as well as shown in Figure 5.51. The bar code replacement of the software jewel case is 'dual purpose'. This provides the usual anti-copy protection and enable Microbar

to function as an authentication device. However, it also contains information which can be cross-referenced with the holders details especially when applied to an identity card, and may provide a link with the product key code on the software which will unlock the loading of the program. Therefore, a copy would, in addition to failing the authenticity test, not provide the same information content and would therefore fail to co-relate with the holders details in case of the ID card and stop the program from loading in the case of the jewel case.

Figure 5.51: Overt Microbar in ID card and software jewel case

Thus the 'Microbar' is more than just an authentication device. The signature is capable of holding data of upto about 15 digits, which, when used in conjunction with an appropriately populated database, can hold a vast amount of information including

- Plant and date of manufacture
- Intended channels of distribution
- Intended countries for sale, etc.

In an era when product diversion is becoming as big a problem as counterfeiting, access to such information, hidden away in a non-changeable medium is becoming a more and more valuable benefit.

5.12 Example of Fractal Texture Watermarking

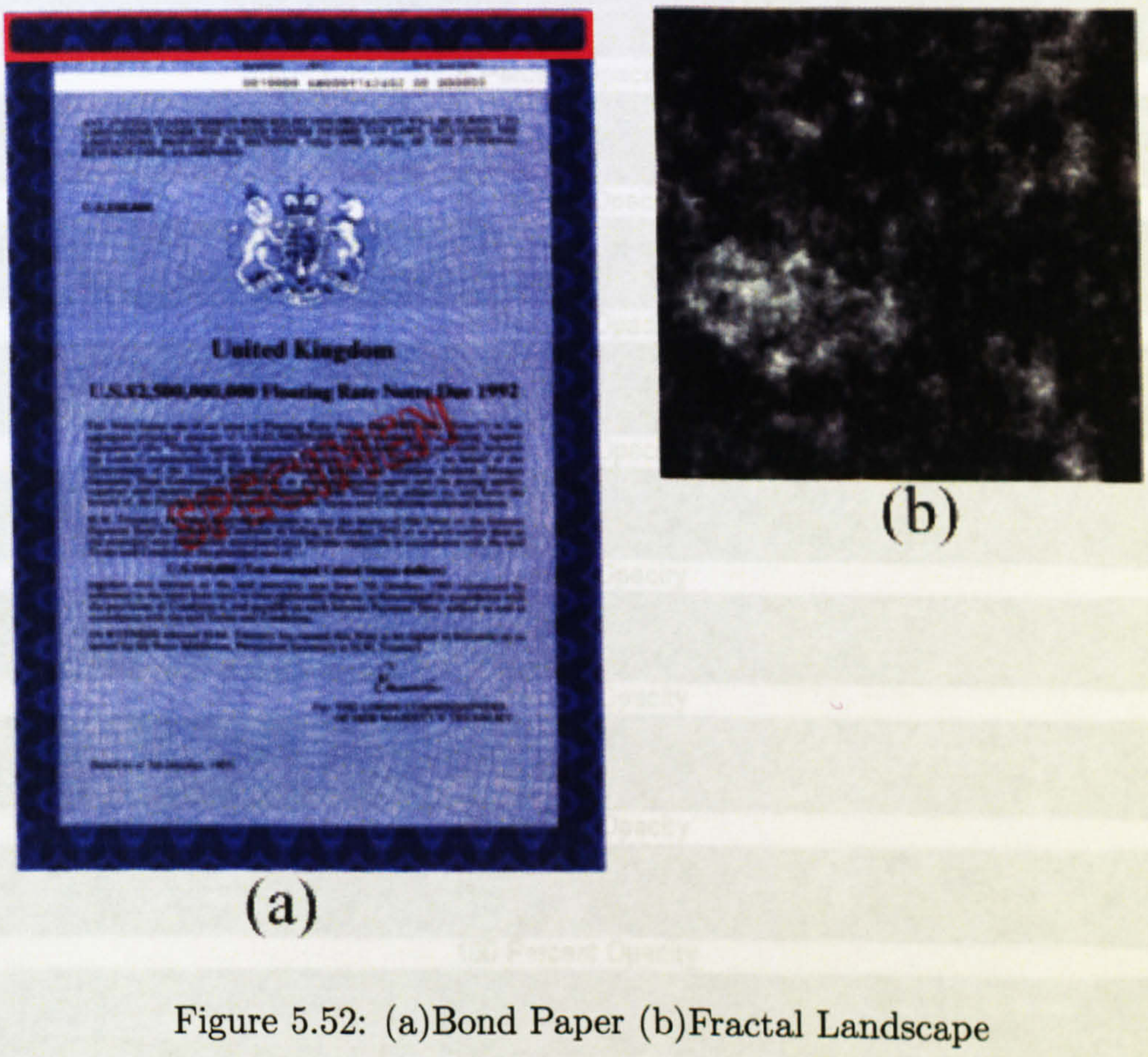


Figure 5.52: (a)Bond Paper (b)Fractal Landscape

Figure 5.52 highlights the process of embedding a fractal landscape into a bond paper. The bond paper (test example) is scanned into Adobe Photo-

shop as shown in Figure 5.52(a). An appropriate fractal landscape is created (Figure 5.52(b)) in MATLAB and is then imported in Adobe Photoshop. The fractal landscape is resized to the exact height and width of a section of the border highlighted in red in Figure 5.52(a) and thereafter embedded in the image. Inspite of the intricate design of the bond paper used in this example, the watermark remains covert, thus proving its robustness.

Figure 5.53: Variation of watermark opacity

Figure 5.53 shows the robustness of the watermark with variation in watermark opacity. The watermark in the figure is created with a spectral exponent, β of 2.4170 and seed random number generator of 117.

Opacity/β	**1.7000**	**2.2000**	**2.7915**	**3.0000**	**3.5000**	**5.0000**
0	0	0	0	0	0	0
5.0000	0.0197	0.0223	0.0261	0.0279	0.0323	0.0413
10.0000	0.1691	0.0872	0.1020	0.1091	0.1262	0.1614
15.0000	0.6172	0.1828	0.2136	0.2283	0.2639	0.3367
20.0000	1.5566	0.3234	0.3775	0.4032	0.4655	0.5920
25.0000	3.1106	0.5019	0.5846	0.6239	0.7187	0.9105
30.0000	5.2453	0.7101	0.8271	0.8824	1.0157	1.2845
50.0000	8.9629	1.8317	2.1211	2.2564	2.5808	3.2222
100.0000	15.3969	6.0376	6.9109	7.3096	8.2480	10.0186

Table 5.5: Variation of watermark opacity and spectral exponent, β

In Table 5.5, spectral exponent β has been varied from 1.7 to 5 for opacity values ranging from 0 to 100 and the MSE has been determined. Again, these values are purely for experimental purposes. Figure 5.54 shows the graphical relationship between these quantities. Noise can be defined by its power spectrum in the form $f^{-\beta}$ as a function of frequency. Most prominent is white noise (noise having a frequency spectrum that is continuous and uniform over a specified frequency band) with spectral exponent, $\beta = 0$. As β increases, the fractal dimension decreases, introducing more noise in the image i.e. for $\beta > 2$, noise changes from white to black (noise that has a frequency spectrum of predominately zero power level over all frequencies except for a few narrow bands or spikes). The graph shows that as the watermark opacity increases, the mean square error also increases linearly. It can be observed that for $\beta = 1.7$, the graph deviates due to brown noise (noise

with brownian motion) present in the image compared to the other values of β considered in the analysis. When the fractal dimension is between two and three, the noise levels are average resulting in low degradation of the image with the watermark present. However, when the fractal dimension is lower than two ($\beta = 5.0$ in this analysis), there is more black noise present resulting in high error. Thus, the fractal texture watermarking technique is a robust scheme as it works fairly well even with high watermark opacity levels which are chosen by the user (in most cases, the maximum being 30%) depending on the application.

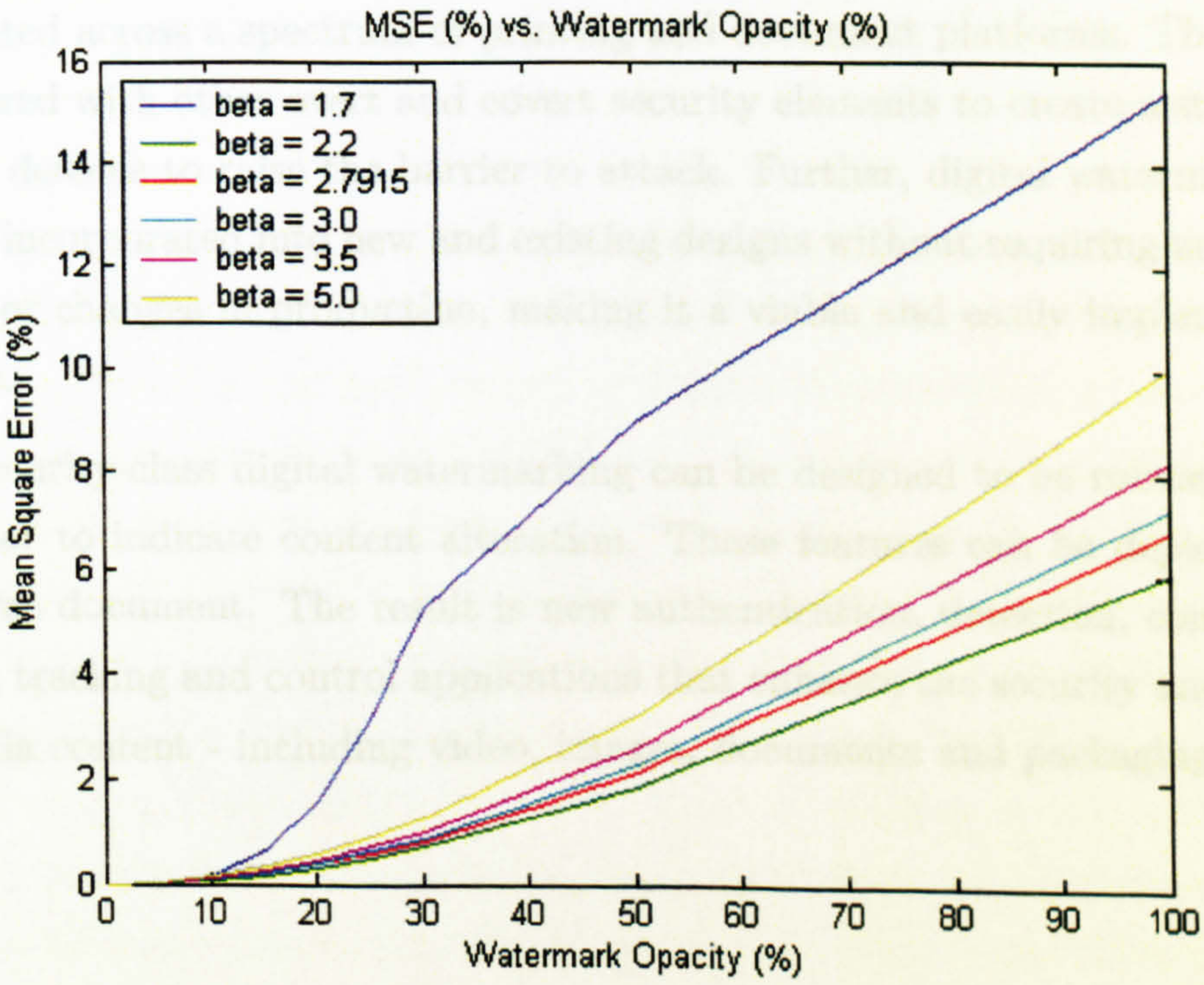


Figure 5.54: MSE versus Watermark Opacity

5.13 Conclusion

Thus, over the past decade, the growth of the digital economy and the increasing popularity of digital imaging products has given counterfeiters and pirates new and powerful tools at affordable prices and fuelled the digital threat to printed and digital content of all kinds.

New digital security tools, such as digital watermarking, have emerged as viable counter measures to digital threats. Today, digital watermarking is deployed in a wide range of systems and has been commercially hardened by leading vendors. Digital watermarking systems have the flexibility to be implemented across a spectrum of printing and document platforms. They can be layered with other overt and covert security elements to create a stronger overall defence to raise the barrier to attack. Further, digital watermarking can be incorporated into new and existing designs without requiring new materials or changes in production, making it a viable and easily implemented feature.

Security-class digital watermarking can be designed to be robust to removal or to indicate content alteration. These features can be deployed in the same document. The result is new authentication, detection, communication, tracking and control applications that enhance the security and value of media content - including video, images, documents and packaging.

Chapter 6

Conclusion and Further Research

After the events of September 11th 2001, much emphasis is being placed on security throughout the world and a lot of research is being carried out in the field of security and applications of Digital Watermarking in areas such as video and audio, internet websites, etc. In the case of the audio watermarking, the main concern however, is that the watermark will be audible and degrade the fidelity of high quality recordings. Myriad techniques for secreting information flow exist; eliminating them is an impossible task. In order to protect ourselves, we must apply the dynamics that serve so well in other unpredictable information security risk venues: assess the risk, find ways to quickly detect the exploit's use, determine an appropriate response and use whatever means available to impede the perpetrator allowing time for detection and reaction mechanisms to work.

However, digital watermarking does have some limitations which have also to be considered. My research presented in this thesis gives an account of the application of Lippmann photography in security and moves on to

fractals as a digital watermark for similar applications.

6.1 Limitations of Digital Watermarking

As of this writing, a counterfeiting scheme has been demonstrated for a class of invertible, feature-based, frequency domain, invisible watermarking algorithms. This counterfeiting scheme could be used to subvert ownership claims because the recovery of the digital signature from a watermarked image requires a comparison with an original[74].

Standard watermarking involves the creation of a watermarked image by encoding a signature into an original image. Authentication proceeds in two stages. First, the watermarks signature is "removed" from the watermarked copy. The watermark signature is the "difference" between the original (white) and the watermarked copy of the original (blue). Next, the extracted signature (blue) is compared against the original signature (gold). Identity signifies authenticity of the copy.

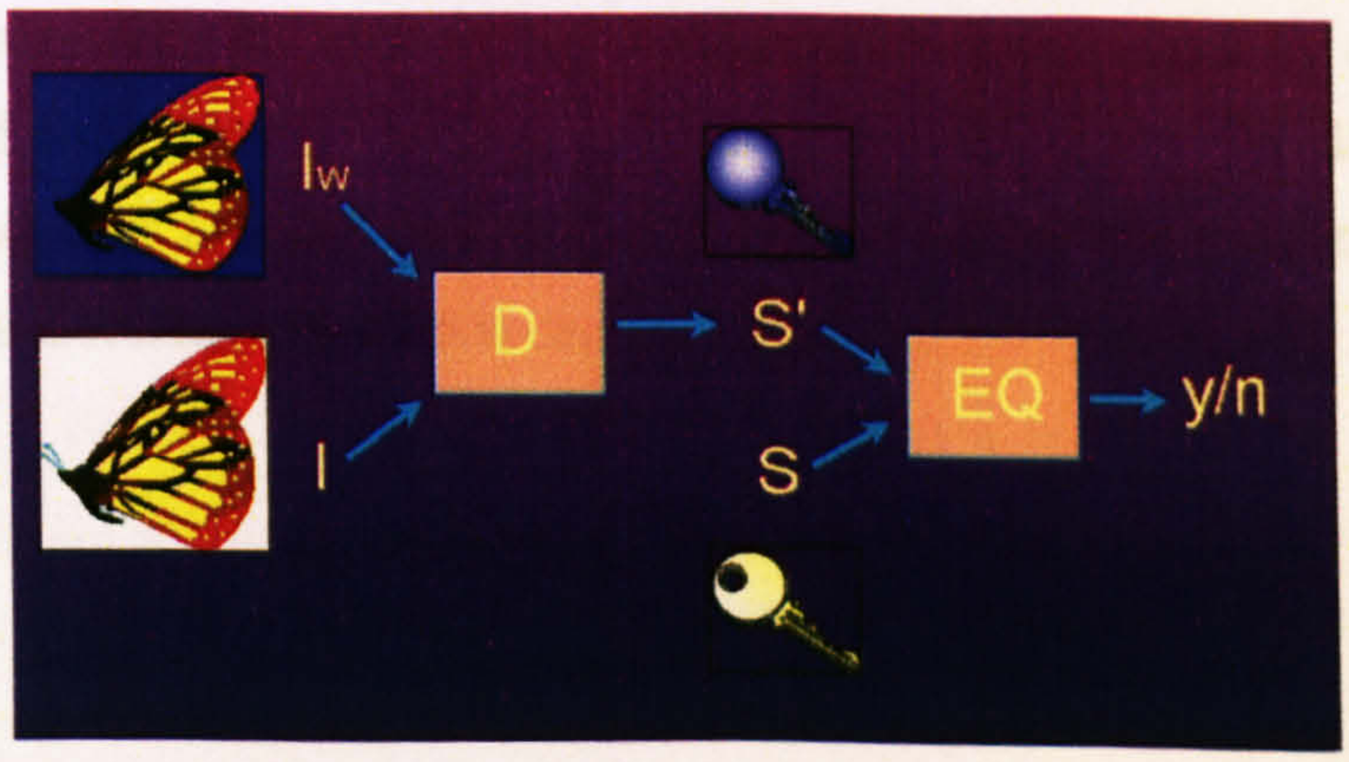


Figure 6.1: Basic Watermarking technique

The counterfeiting scheme (Figures 6.1 and 6.2) works by first creating a counterfeit watermarked copy (violet) from the genuine watermarked copy (blue) by effectively inverting the genuine watermark. This inversion produces a counterfeit signature (violet) as well.

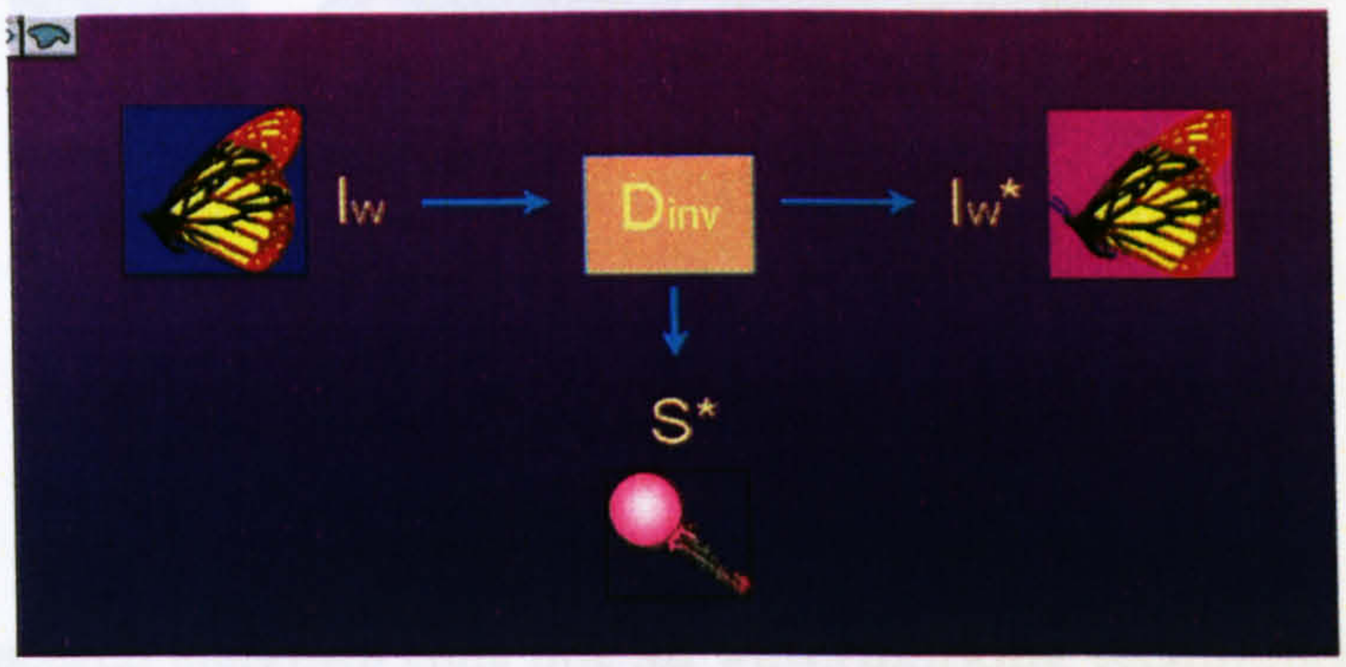


Figure 6.2: Watermark "inversion" for counterfeiting

The trick is that the original image and bonafide signature stand in the same relationship to the watermarked image as the counterfeit image and counterfeit signature (Figure 6.3). Thus, the technique of establishing legitimate ownership by recovering the signature watermark by comparing a watermarked image with the original image breaks down. While it may be demonstrated that at least one recipient has a counterfeit watermarked copy, it can not be determined who it is.

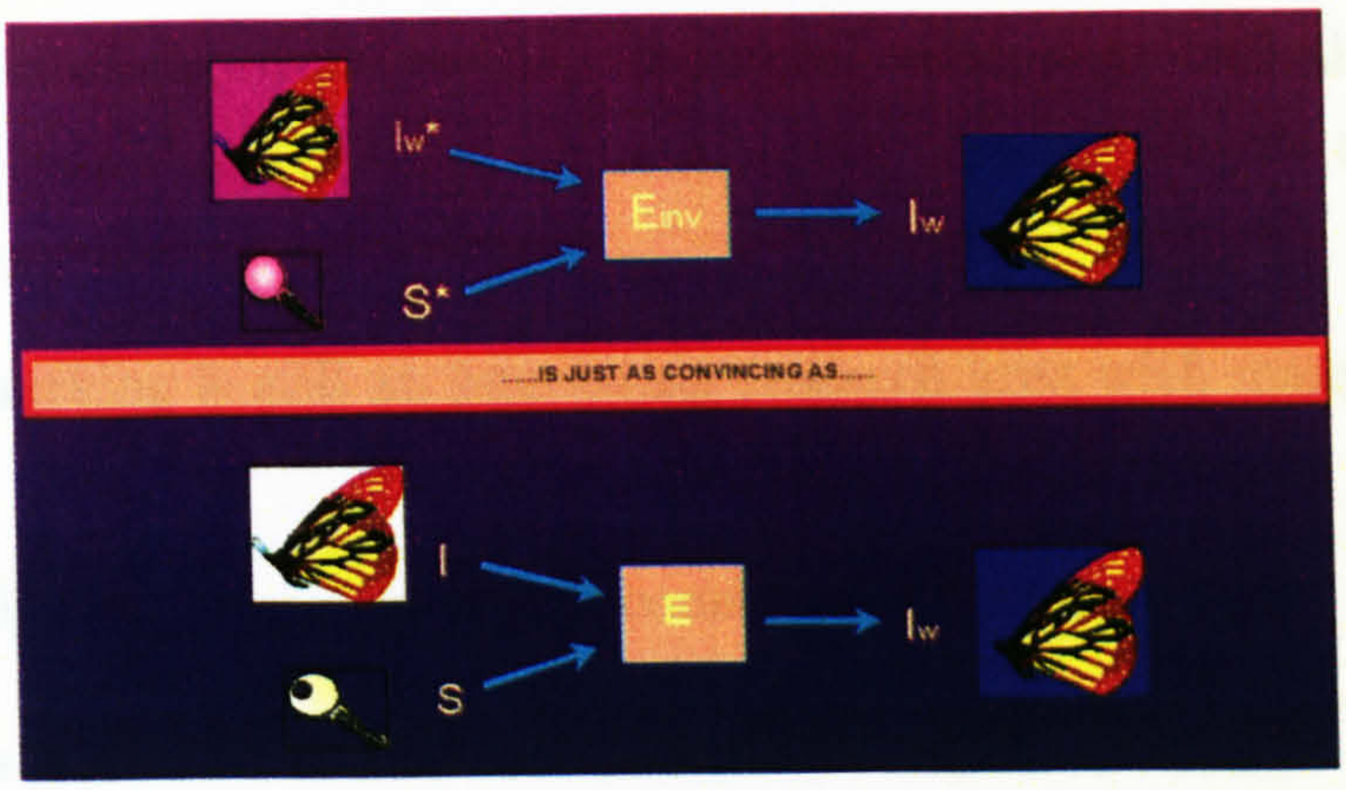


Figure 6.3: Counterfeit logic

This research suggests that not all watermarking techniques will be useful in resolving ownership disputes in courts of law. There will likely be non-commercial applications, or those with limited vulnerability to theft, where "good enough watermarking" will suffice. More sensitive applications may require non-invertable or non-extracting watermarking techniques. These issues are under consideration at the present time.

6.2 The Future of Digital Watermarking

The enormous popularity of the World Wide Web in the early 1990's demonstrated the commercial potential of offering multimedia resources through the digital networks. Since commercial interests seek to use the digital networks to offer digital media for profit, they have a strong interest in protecting their ownership rights. Digital watermarking has been proposed as one way to protect such interests. Though much research remains before watermarking systems become robust and widely available, there is much promise that they will contribute significantly to the protection of propri-

etary interests of electronic media. Collateral technology will also be necessary to automate the process of authentication, non-repudiable transmission and validation.

Watermarking is still an interesting research area with many interesting problems.

The following issues are worth considering for the future:

- Where will it be useful?
- Will watermarking only be used as a second-tier security system?
- Will there be significant theoretical developments?
- Legal and political issues:
  - Watermarking technologies have not been tested in court
  - Is Watermarking the 'feel-good' technology of multimedia?

6.2.1 Video Watermarking

A video sequence cannot simply be treated as an ordered collection of images:

- Visibility issues in the use of 'still' image watermarks
- Visibility issues in stop frames
- Human perception of motion is not accounted for in visual models for still images
- Embedding the same watermark in all the frames of a video sequence is not secure, an attacker can correlate across the entire sequence to estimate the watermark (temporal collusion)

- Embedding completely different watermarks in successive frames of a video sequence is not secure
- Successive video frames are highly correlated, an attacker can exploit this estimate and remove a watermark
- The techniques for compressing video do not necessarily encode each frame of the sequence identically
- The synchronization of the audio with the video sequence may be a consideration for watermark protection
- Video Watermarking can use still image approaches but may have problem in MPEG with B and P frames
- Hash parts of the compressed video stream
- Techniques could be used to prevent multiple viewing, copying and editing (e.g. inserts)
- Another point to consider is can the watermark survive the conversion back to an analog signal?
- Video watermarking cannot trivially extend image watermarking techniques, additional attacks are possible and computationally very expensive
- Unique attacks on video watermarks such as frame shuffling/insertion and inter-frame collusion should be investigated

6.2.2 Two-Dimensional Digital Watermark

An extensive review on existing two-dimensional watermarking schemes has been presented in Chapter Three. A lot of research is being done on

various methods to extend the one-dimensional watermarking to the second dimension. Methods such as FFT (Fast Fourier Transform), DCT (Discrete Cosine Transform), DFT (Discrete Fourier Transform), Wavelets, Linear Predictive Coding, Lapped Orthogonal Transforms, Fractals are used to embed the watermark. Some watermarking schemes are based on Fractal Compression Algorithms i.e. on the resolution of the inverse problem for Iterated Function Systems (IFS).

6.2.3 Three-Dimensional Digital Watermark

At the Stanford University, USA, research is being carried out about the possible use of three-dimensional watermarking in artifacts and collectibles. The Digital Michelangelo Project involved digitizing the sculptures and architecture of Michelangelo. The three-dimensional watermark would be introduced in the computer model of the artefacts to prevent the replication. For further information, please refer to [112].

6.3 Conclusion

Given the nature of the technology and the scope of the business issues that are affected, it is difficult to predict with any certainty where we will be in 5 years time. Nevertheless, the following predictions are presented as a reasonable set of possibilities:

- Watermarking techniques and tools will become more common and increasingly sophisticated.

- A stego process will be developed to embed trojans, worms and viruses in media such as images or audio files and have them become active by

viewing or listening to the files. In 2001, the Nimda worm demonstrated that it was possible to get a virus just by visiting an infected web site. In January of 2002, viruses were being delivered by Macromedia flash images. One day, merely viewing a bitmap image might cause a virus attack on your PC.

- Intrusion Detection Systems (IDS) will include images as part of their attack signatures.
- Anti-virus software will be developed with steganalytical capabilities to detect viruses in audio and image files.
- A strong tamper-resistant, economically viable digital watermark will be developed.

The person who develops the last item will undoubtedly become very rich. There is probably no better motivating factor for learning about Digital Watermarking!

Appendix A

Mathematical Modelling of the Lippmann Process; Work done by Nareid and Pedersen[107]

The first researchers to perform computer simulation of the Lippmann process were Nareid and Pedersen[107]. For monochromatic recording, Kogelnik's coupled-wave theory can be applied to Lippmann photography. For polychromatic recording, superposition of several frequencies in the recording light gives rise to aperiodic space gratings for which the Kogelnik theory cannot be applied. Instead, Nareid and Pedersen based their modelling on the theory of wave propagation in a stratified medium combined with first Born approximation. Fournier and Burnett investigated old Lippmann photographs using modern electron microscopy. They recorded a monochromatic volume grating in DuPont photopolymer material using filtered light (520nm, 10nm bandwidth) from a slide projector. Recently, Rich and Dickerson presented a report on recording colour photographs on both Slavich PFG-03c plates as well as emulsions prepared by them according to old Lippmann methods.

Because of the difficulties in obtaining isochromatic plates, a combination of the interference recording method and the three-colour technique was proposed by Lippmann. By making a sequential recording through three colour filters (blue, green and red) it was possible to individually control the exposure times through the three filters, thus obtaining good colour balance. Given a sufficient emulsion thickness and the correct processing methods, very pure spectrum colours can be obtained[89].

Mathematical Modelling done by Nareid and Pedersen

In the computer modelling of the Lippmann process spectral response, we discretize the model of the developed emulsion by subdividing it into a large number of homogenous layers with constant refraction index[107]. The layer thickness is chosen freely, and it is found that no appreciable quantization noise is observed with a layer thickness less than $\lambda/20$. Then the entire emulsion is modelled by a stack of 500-1000 layers (Temporal Coherence Function).

For each layer the wave equation is solved exactly, and the solution is matched to those of the adjacent layers. This leads to the well-known exact matrix formulation of wave propagation in stratified media. Each layer is represented by a characteristic matrix

$$M(\Delta) = \begin{bmatrix} \cos\delta & -\frac{i}{p}\sin\delta \\ -ip\sin\delta & \cos\delta \end{bmatrix} \tag{A.1}$$

where Δ is the layer thickness and for normal incidence, p=N is the complex refractive index of the layer, while $\delta = kN\Delta$ is the complex phase delay of the layer. For a sequence of layers, the total characteristic matrix is now

simply given as the product of those of the individual layers, i.e.;

$$M_{total} = \begin{bmatrix} m11 & m12 \\ m21 & m22 \end{bmatrix} = \prod M(\Delta) \tag{A.2}$$

The reflection and transmission coefficients of the emulsion are now found by the formulae

$$R = \frac{(m11 + m12pl)p1 - (m21 + m22pl)}{(m11 + m12pl)p1 + (m21 + m22pl)} \tag{A.3}$$

$$T = \frac{2p1}{(m11 + m12pl)p1 + (m21 + m22pl)} \tag{A.4}$$

Here $p1$ and pl refer to the first and the last layers respectively. The reflectivity and transmissivity are then found from

$$R = |R|^2 \qquad T = \frac{pl}{p1}|T|^2 \tag{A.5}$$

The complex refractive index is computed for each of the 500-1000 layers and stored in an array. The spectral response is then computed by applying the matrix solution to the entire stack of layers for each of the frequencies in question. Further results are shown overleaf.

The Figures A.1 (a),(b),(c),(d) show the spectral response of a single frequency recording with linear index modulation of different modulation contrast in the loss less case.

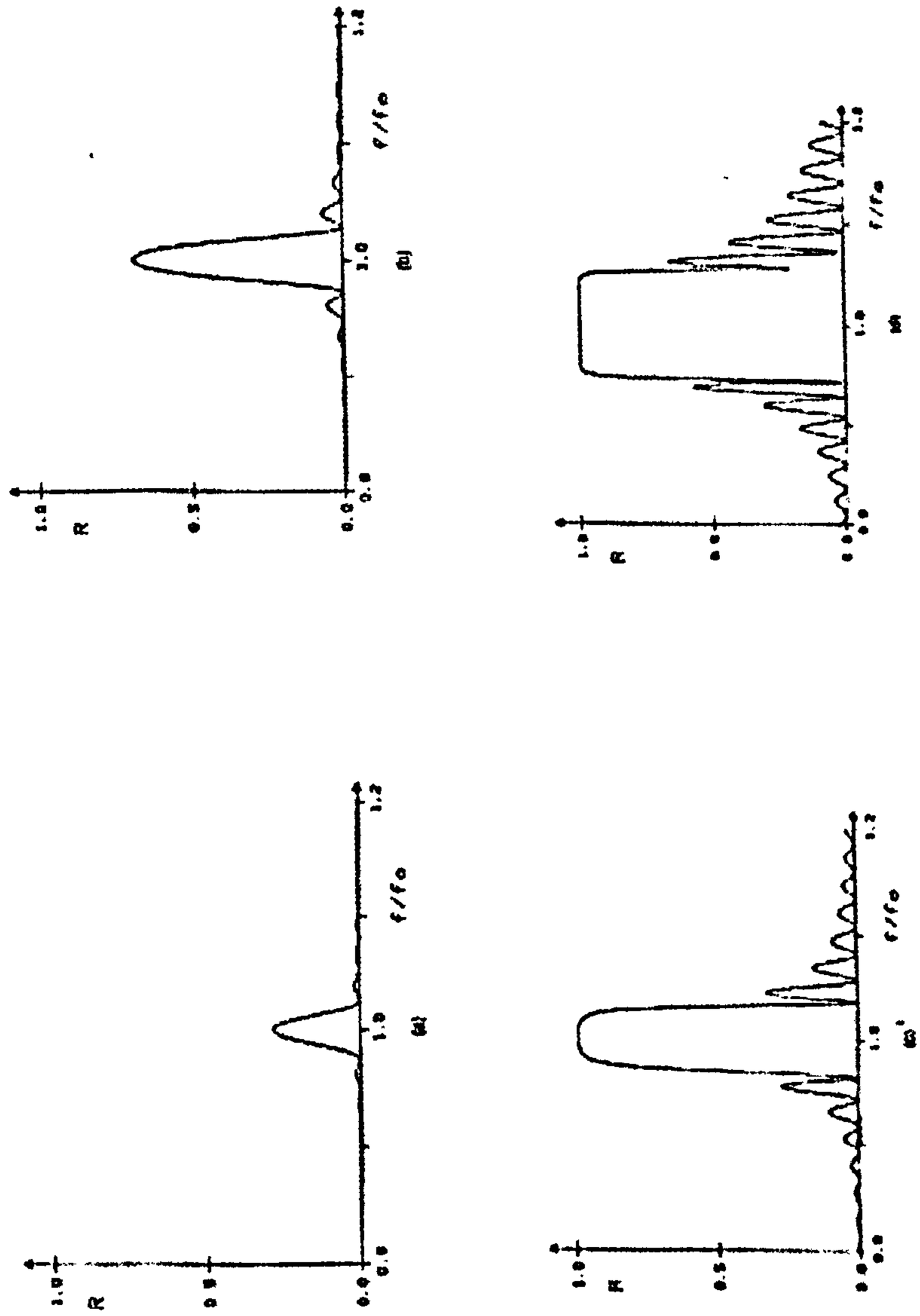


Figure A.1: Spectral response of single-frequency recording in the lossless case for (a)μ_n=0.01, (b)μ_n=0.02, (c)μ_n=0.05, (d)μ_n=0.10

As the modulation contrast is increased, the pattern is amplified until the central peak is saturated i.e. the reflectivity reaches R=1. This causes a

broadening of the central peak, while the secondary peaks are further amplified. The broadening of the central peak is a Bragg-resonance phenomenon. As the index modulation increases, a larger part of the incident wave energy is reflected at each modulation period, and, as a result, the incident wave will be extinguished before it extends through the emulsion. Therefore the broadening of the central peak can be seen as the result of a reduced effective depth of the scattering volume.

The Figures A.2(a) and (b) show single-frequency recordings with loss. The dashed lines refer to the loss less curves for comparison purposes. The above single frequency results are in close agreement with the well-known results of Kogelnik's coupled wave analysis.

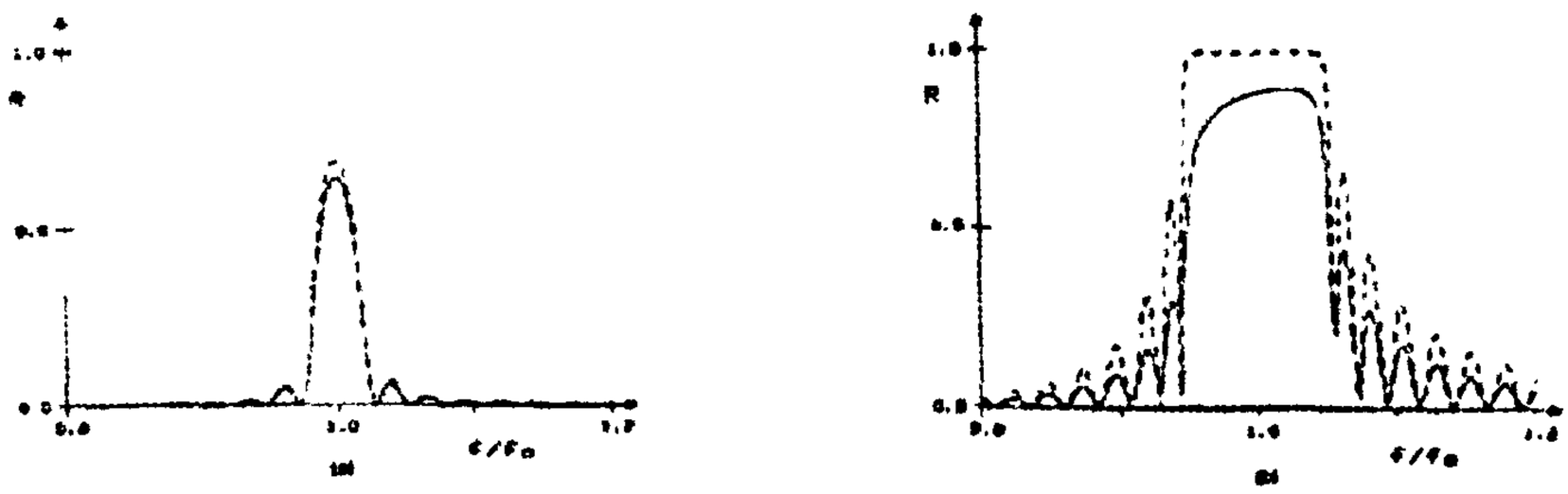


Figure A.2: Spectral response of single-frequency recordings with loss

The Figure A.3 shows the result of a wideband recording in the loss less case. These curves all show strong spectral distortion relative to the rectangular input spectrum. This distortion is due to the fact that only one side of the temporal coherence function is being recorded in the emulsion.

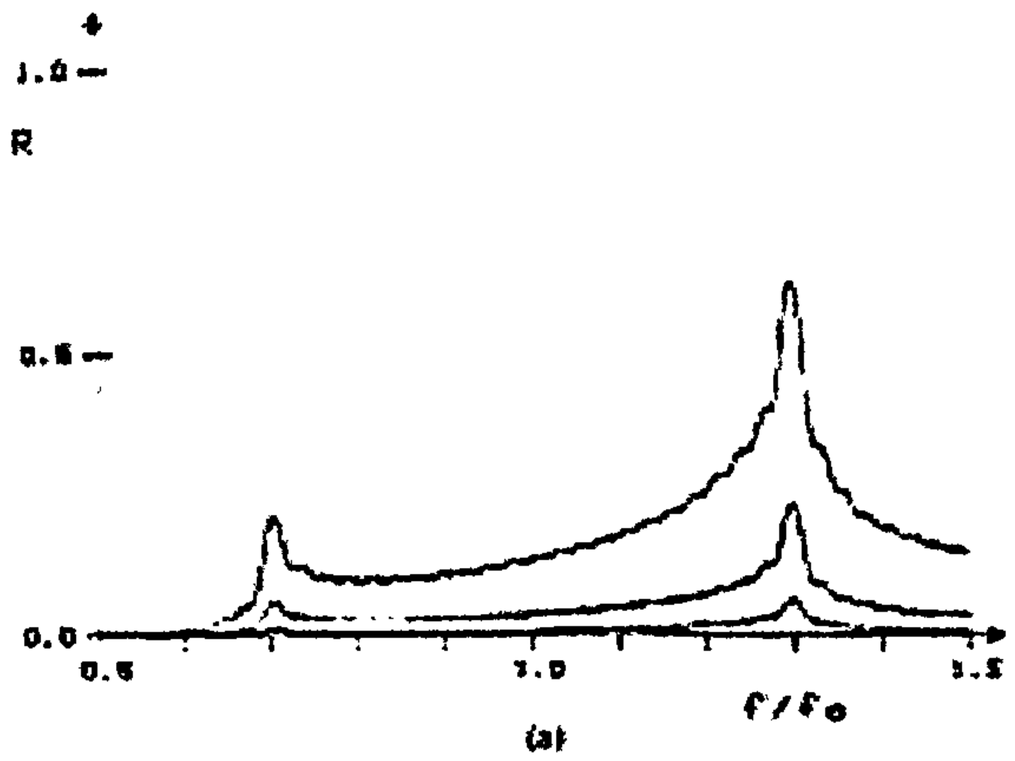


Figure A.3: Spectral response for wideband recording in the lossless case

The Figure A.4 shows a wideband recording with 50% absorption.

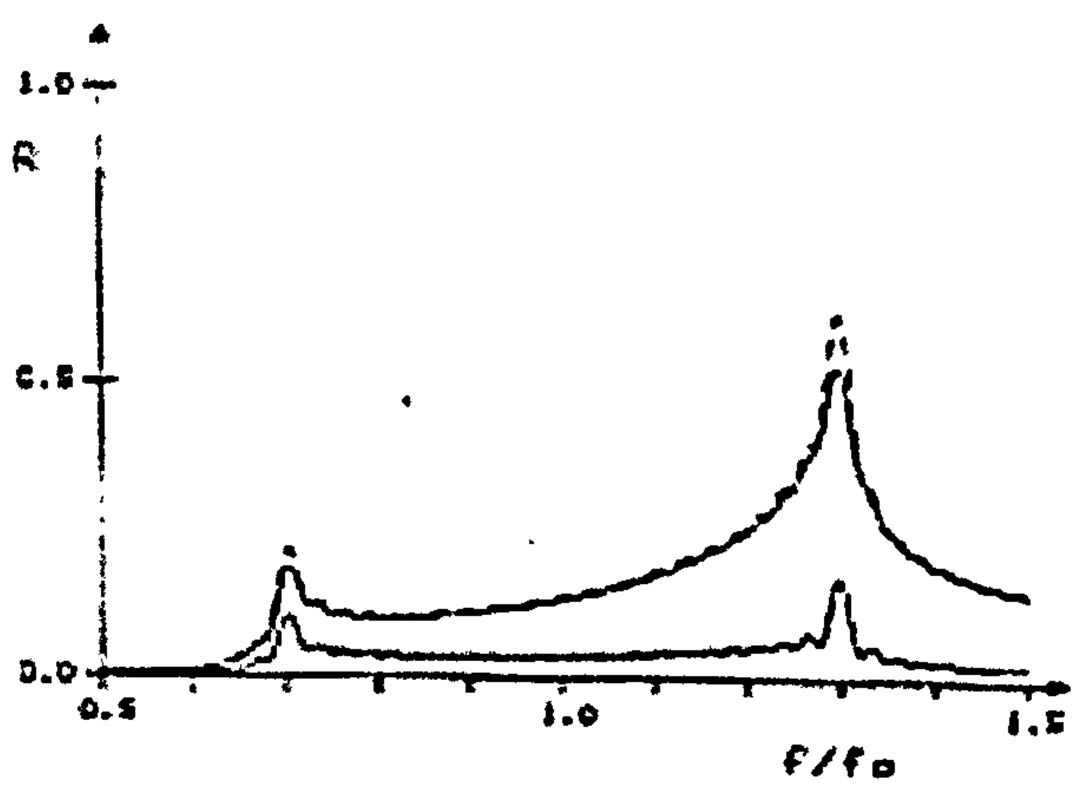


Figure A.4: Spectral response of wideband recording with 50% loss

The results are similar to the earlier results with an interesting difference. In the loss less case we have reciprocity i.e. the reflectivity is the same if the emulsion is illuminated from the other side, but with absorption this reciprocity is lost. The upper curve shows the result of ordinary reconstruction i.e. illumination from the front side of the emulsion, whereas the lower

curve shows the corresponding result for illumination from the backside of the emulsion.

The effect of film nonlinearity is tested by the extreme case of a hard-limiting emulsion in which the refractive index is set equal to the cut off level if it exceeds that level. The results are shown in the Figures A.5(a), (b), (c) and (d).

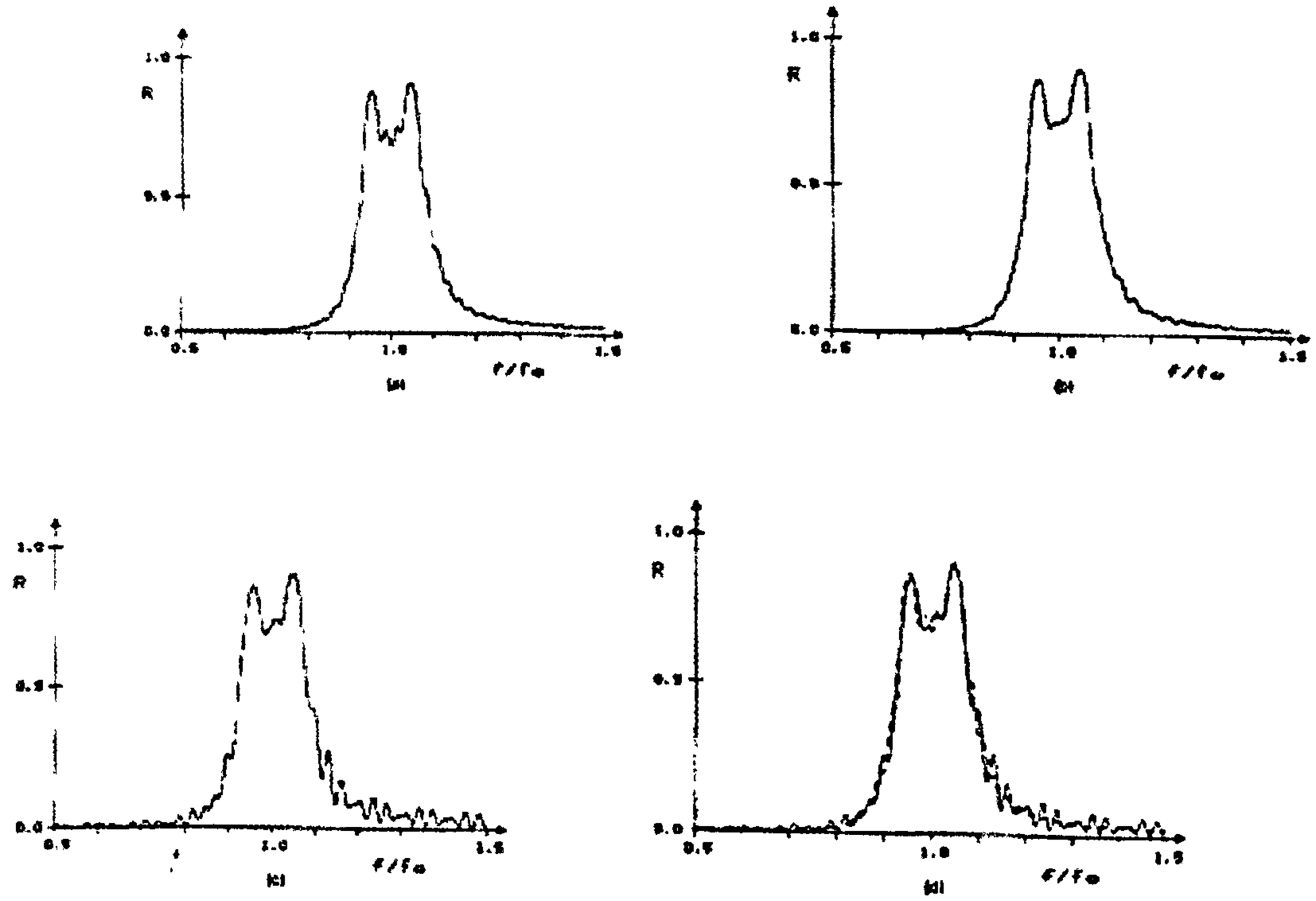


Figure A.5: Illustration of the effects of film nonlinearity; (a)no clipping, (b)50% clipping, (c)75% clipping, (d)comparison between the results with 75% clipping and with no clipping

This shows that the process is virtually insensitive to film nonlinearities. This is because nonlinearities like clipping introduce higher harmonics without causing serious deterioration in the first harmonic of the spatial index modulation signal. These higher harmonics have Bragg resonances at multiples of the mean recording frequency and will not contribute to the spectral response in the visible part of the spectrum.

Appendix B

Optical Diffraction Theory and Image Formation[63]

In Chapter 4, a model for generating speckle was demonstrated based on lowpass filtering white Gaussian noise. This 'standard approach' for producing a basic texture that occurs in nearly all coherent imaging systems is based on a convolution model for the image - the convolution of the object function with the point spread function of the imaging system. This model is of fundamental importance to all imaging systems and understanding where it comes from is therefore necessary if the reader wishes to appreciate the physics of imaging systems in general. This is one of the underlying themes of what may be called image understanding and requires some advanced analysis associated with the propagation and scattering of wavefields. This appendix has been written to help those readers who are interested come to terms with the theory of image formation, which is fundamental to image understanding. Starting from first principles (i.e. by solving a well known wave equation), we take the reader through a calculation in which the analytical techniques used are fundamental to modelling most imaging systems. The emphasis is on optical imaging and optical image formation and the principles of what is

commonly known as 'Fourier optics'. However, many of the concepts, ideas and analytical techniques can be applied to other imaging modalities. The whole aim of this analysis is to quantify coherent and incoherent (optical) image formation and thus relate some fundamental results of physical optics to the digital imaging models used upon which fractal image analysis, like any other form of image analysis and processing, ultimately depends.

We start by investigating the scalar theory of optical diffraction and develop the so called Kirchhoff diffraction theory. This leads directly to two types of optical diffraction models known as Fraunhofer and Fresnel diffraction. In turn, the former model for diffraction leads us to investigate the Fourier transforming properties of a lens and consider the principles of Fourier optics (the Abbe theory of imaging) in which a lens (in the focal plane) can be taken to perform a Fourier transform of the object that is being focused. This principle leads directly to the basic convolution model for an optical imaging system and the concept of optical image formation in general from which specific models for coherent and incoherent image formation can be derived.

Scalar Diffraction Theory

The phenomenon of diffraction is common to 1D transverse (and hence scalar) waves such as water waves, true scalar waves such as in acoustics and vector waves as in optics. All three cases appear to exhibit the same magnitude of diffraction phenomena.

In order to approximately describe optical diffraction, it is reasonable to first adopt a scalar model for light. This type of model is concerned with

a monochromatic scalar 'disturbance'

$$V(\mathbf{r},t) = U(\mathbf{r})\exp(-i\omega t)$$

where U is the scalar complex amplitude, ω is the angular frequency (=$2\pi\times$ frequency), t is time and $\mathbf{r}$ is a three dimensional vector which in Cartesian coordinates is given by

$$\mathbf{r} = \mathbf{x}x + \mathbf{y}y + \mathbf{z}z$$

In free space V satisfies the homogeneous wave equation

$$\nabla^2 V - \frac{1}{c^2}\frac{\partial^2 V}{\partial t^2} = 0$$

where c is the velocity of light and U satisfies the homogeneous Helmholtz equation

$$\nabla^2 U + k^2 U = 0$$

where

$$k = \frac{\omega}{c} = \frac{2\pi}{\lambda}$$

Here, k is the wavenumber and λ is the wavelength.

Scalar diffraction theory (which essentially stems from solutions to the above equation - subject to careful interpretation of the physical significance of such solutions) should be regarded as a first approximation to optical diffraction.

The observed intensity I (the observed quantity at optical frequencies) can be taken to be given by (by definition)

$$I = |\, U \,|^2$$

Except in free space, U is not (in general) a Cartesian component of the vector electric or magnetic field. Scalar diffraction theory is accurate if:

(i) The diffracting aperture is large compared to the wavelength.

(ii) The diffracted fields are observed at a reasonable distance from the screen.

Kirchhoff Diffraction Theory

Consider a scalar wave field U described by the homogeneous Helmholtz equation

$$(\nabla^2 + k^2)U = 0$$

Let U_i be the field incident on a surface S and introduce the following (Kirchhoff) boundary conditions

$$U = U_i, \quad \frac{\partial U}{\partial \hat{\mathbf{n}}} = \frac{\partial U_i}{\partial \hat{\mathbf{n}}} \text{ on S}$$

In the case of an aperture formed by constructing a hole in a screen, the conditions above apply over the plane of the aperture and U_i may be considered to be the field incident on the plane of the screen in the absence of the screen itself.

The Kirchhoff boundary conditions ignore edge effects at an aperture and thus will only be valid for apertures very much smaller than a wavelength. The diffraction theory which stems from a solution to the Helmholtz equation based on these boundary conditions is called the Kirchhoff diffraction theory.

Green's function solution

Consider the Green's function G which is the solution to

$$(\nabla^2 + k^2)G = -\delta(\mathbf{r} - \mathbf{r_0})$$

and given by

$$G(\mathbf{r} \mid \mathbf{r_0}, k) = \frac{1}{4\pi \mid \mathbf{r} - \mathbf{r_0} \mid} \exp(ik \mid \mathbf{r} - \mathbf{r_0} \mid)$$

We can construct two equations:

$$G\nabla^2 U + k^2 GU = 0$$

$$U\nabla^2 G + k^2 UG = -U\delta$$

Subtracting these equations and integrating over a volume V we obtain

$$\iiint_V U\delta dV = \iiint_V (G\nabla^2 U - U\nabla^2 G)\, dV$$

Using the shifting property of the delta function together with Green's theorem, we obtain a solution for the field U at $\mathbf{r_0}$, i.e.

$$U(\mathbf{r_0}) = \iint_S \left(G\frac{\partial U}{\partial \hat{\mathbf{n}}} - U\frac{\partial G}{\partial \hat{\mathbf{n}}} \right) dS$$

Introducing the Kirchhoff boundary conditions we arrive at the basic Kirchhoff diffraction formula given by

$$U(\mathbf{r_0}) = \iint_S \left(G\frac{\partial U_i}{\partial \hat{\mathbf{n}}} - U_i\frac{\partial G}{\partial \hat{\mathbf{n}}} \right) dS$$

This equation is sometimes called the Kirchhoff integral. To compute the diffracted field using the Kirchhoff integral an expression for U_i must be introduced and the derivatives $\partial/\partial\hat{\mathbf{n}}$ with respect to U_i and G computed.

Consider the case where the incident field is a plane wave field of unit amplitude (with wavenumber $k \equiv |\,\mathbf{k}\,|$, $\hat{\mathbf{k}} = \mathbf{k}/k$). Then

$$U_i = \exp(i\mathbf{k}\cdot\mathbf{r})$$

and

$$\frac{\partial U_i}{\partial \hat{\mathbf{n}}} = \hat{\mathbf{n}}\cdot\nabla \exp(i\mathbf{k}\cdot\mathbf{r}) = i\mathbf{k}\cdot\hat{\mathbf{n}} \exp(i\mathbf{k}\cdot\mathbf{r}) = ik\hat{\mathbf{n}}\cdot\hat{\mathbf{k}} \exp(i\mathbf{k}\cdot\mathbf{r})$$

The calculation of $\partial G/\partial \hat{\mathbf{n}}$ is more complicated.

$$\frac{\partial G}{\partial \hat{\mathbf{n}}} = \hat{\mathbf{n}} \cdot \nabla G$$

and

$$\begin{aligned}
\nabla G &= \mathbf{x}\frac{\partial}{\partial x}\frac{\exp(ik\sqrt{(x-x_0)^2+\cdots})}{4\pi\sqrt{(x-x_0)^2+\cdots}} + \cdots \\
&= -\mathbf{x}\frac{1}{4\pi}\exp(ik\sqrt{(x-x_0)^2+\cdots})[(x-x_0)^2+\cdots]^{-\frac{3}{2}}(x-x_0) \\
&\quad +\mathbf{x}\frac{ik}{4\pi}\frac{(x-x_0)}{(x-x_0)^2+\cdots}\exp(ik\sqrt{(x-x_0)^2+\cdots}) + \cdots \\
&= \mathbf{x}\frac{\exp(ik\sqrt{(x-x_0)^2+\cdots})}{4\pi\sqrt{(x-x_0)^2+\cdots}}\frac{(x-x_0)}{\sqrt{(x-x_0)^2+\cdots}}\left(ik - \frac{1}{\sqrt{(x-x_0)^2+\cdots}}\right) + \cdots \\
&= \hat{\mathbf{m}}\left(ik - \frac{1}{|\,\mathbf{r}-\mathbf{r_0}\,|}\right)G
\end{aligned}$$

where

$$\hat{\mathbf{m}} = \frac{\mathbf{r}-\mathbf{r_0}}{|\,\mathbf{r}-\mathbf{r_0}\,|}$$

Therefore,

$$\frac{\partial G}{\partial \hat{\mathbf{n}}} = \hat{\mathbf{n}} \cdot \hat{\mathbf{m}}\left(ik - \frac{1}{|\,\mathbf{r}-\mathbf{r_0}\,|}\right)G$$

In most practical circumstances the diffracted field is observed at distances $|\,\mathbf{r}-\mathbf{r_0}\,|$ where

$$|\,\mathbf{r}-\mathbf{r_0}\,| >> \lambda$$

This condition allows us to introduce the simplification

$$\nabla G \simeq ik\hat{\mathbf{m}}G$$

so that

$$\frac{\partial G}{\partial \hat{\mathbf{n}}} \simeq ik\hat{\mathbf{n}} \cdot \hat{\mathbf{m}}G$$

The Kirchhoff diffraction formula then reduces to the form

$$U(\mathbf{r_0}) = ik\iint_S \exp(i\mathbf{k}\cdot\mathbf{r})(\hat{\mathbf{n}}\cdot\hat{\mathbf{k}} - \hat{\mathbf{n}}\cdot\hat{\mathbf{m}})GdS$$

Fraunhofer Diffraction

Fraunhofer diffraction assumes that the diffracted wavefield is observed a large distance from the screen. The point of observation is in the far field. For this reason, Fraunhofer diffraction is sometimes called diffraction in the 'far field'. Mathematically, it represent an asymptotic solution to the problem posed by the Kirchhoff diffraction formula given above.

The basic idea is to exploit the simplifications that can be made to the Kirchhoff diffraction integral by considering the case when

$$|\,\mathbf{r}\,| << |\,\mathbf{r_0}\,|$$

As the point $\mathbf{r}$ moves in the domain of integration, the complex exponent describing the Green's function changes rapidly but if $r_0 >> r$ where $r_0 \equiv |\,\mathbf{r_0}\,|$ and $r \equiv |\,\mathbf{r}\,|$ then

$$\frac{1}{|\,\mathbf{r}-\mathbf{r_0}\,|} \simeq \frac{1}{r_0}$$

and

$$\hat{\mathbf{n}}\cdot\hat{\mathbf{k}} - \hat{\mathbf{n}}\cdot\hat{\mathbf{m}} \simeq \hat{\mathbf{n}}\cdot\hat{\mathbf{k}} + \hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_0$$

where

$$\hat{\mathbf{r}}_0 = \frac{\mathbf{r_0}}{r_0}$$

In this case, the Kirchhoff diffraction integral reduces to

$$U(\mathbf{r_0}) \simeq \frac{ik\alpha}{4\pi r_0}\iint_S \exp(i\mathbf{k}\cdot\mathbf{r})\exp(ik\,|\,\mathbf{r}-\mathbf{r_0}\,|)dS$$

where

$$\alpha = \hat{\mathbf{n}}\cdot\mathbf{k} + \hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_0$$

The next simplification that can be made under the condition $r_0 >> r$ is to the exponent

$$\exp(ik\,|\,\mathbf{r}-\mathbf{r_0}\,|)$$

Now,

$$\begin{aligned} |\,\mathbf{r}-\mathbf{r}_0\,| &= [(\mathbf{r}-\mathbf{r}_0)\cdot(\mathbf{r}-\mathbf{r}_0)]^{\frac{1}{2}} \\ &= [r^2-2\mathbf{r}\cdot\mathbf{r}_0+r_0^2] \\ &= r_0\left(1-2\frac{\mathbf{r}\cdot\mathbf{r}_0}{r_0^2}+\frac{r^2}{r_0^2}\right)^{\frac{1}{2}} \end{aligned}$$

Introducing a binomial expansion of this result,

$$\begin{aligned} |\,\mathbf{r}-\mathbf{r}_0\,| &= r_0-\mathbf{r}\cdot\hat{\mathbf{r}}_0+\frac{r^2}{2r_0}+\cdots \\ &\simeq r_0-\mathbf{r}\cdot\hat{\mathbf{r}}_0;\quad r<<r_0 \end{aligned}$$

the Kirchhoff diffraction integral reduces to

$$U(\mathbf{r}_0)\simeq\frac{ik\alpha}{4\pi r_0}\exp(ikr_0)\iint_S\exp(i\mathbf{k}\cdot\mathbf{r})\exp(-ik\hat{\mathbf{r}}_0\cdot\mathbf{r})dS$$

We are now in a position to introduce the geometry of the 'aperture system' which can be described using the Cartesian coordinates:

$$\mathbf{r}=\hat{\mathbf{x}}x+\hat{\mathbf{y}}y+\hat{\mathbf{z}}z$$

$$\mathbf{r}_0=\hat{\mathbf{x}}x_0+\hat{\mathbf{y}}y_0+\hat{\mathbf{z}}z_0$$

Here, x_0 and y_0 can be taken to describe the position on a flat screen at a distance z_0 from the diffracting aperture. Consider the following physical conditions:

$\hat{\mathbf{k}}=\hat{\mathbf{z}}$	The aperture is illuminated by a plane wave at normal incidence
$\hat{\mathbf{r}}_0\simeq\hat{\mathbf{z}}$	The diffraction pattern is observed only at small angles
$kz\rightarrow 0$	The aperture is 'infinitely thin'

The first two conditions give

$$\alpha\simeq\hat{\mathbf{n}}\cdot\hat{\mathbf{k}}+\hat{\mathbf{n}}\cdot\hat{\mathbf{r}}_0\simeq 2$$

and with the third condition we obtain

$$U(x_0, y_0, z_0) = \frac{i}{\lambda}\frac{\exp(ikr_0)}{r_0}\iint_S \exp\left[-\frac{ik}{r_0}(xx_0 + yy_0)\right] dxdy$$

This equation gives the amplitude at (x_0, y_0, z_0) in the far field when the aperture is illuminated by a plane wave at normal incidence.

Since a point of observation lies in a plane (the observation screen) located at a fixed distance z_0 from the aperture,

$$\begin{aligned} r_0 &= \sqrt{x_0^2 + y_0^2 + z_0^2} \\ &= z_0\left(1 + \frac{x_0^2 + y_0^2}{z_0^2}\right)^{\frac{1}{2}} \\ &\simeq z_0 + \frac{x_0^2 + y_0^2}{2z_0} \end{aligned}$$

Using this expression for r_0 in the exponent $\exp(ikr_0)$ but using $r_0 \simeq z_0$ elsewhere, we have

$$U(x_0, y_0) = \frac{i}{\lambda}\frac{\exp(ikz_0)}{z_0}\exp\left(ik\frac{x_0^2 + y_0^2}{2z_0}\right)\iint_S \exp\left(-\frac{ik}{z_0}(xx_0 + yy_0)\right) dxdy$$

Finally, let the aperture be filled with an arbitrary distribution $f(x, y)$ that is zero outside S, then

$$U(x_0, y_0) = \frac{i}{\lambda}\frac{\exp(ikz_0)}{z_0}\exp\left(ik\frac{x_0^2 + y_0^2}{2z_0}\right)\int_{-\infty}^{\infty} f(x, y)\exp\left(-\frac{ik}{z_0}(xx_0 + yy_0)\right) dxdy$$

where U is the amplitude of the Fraunhofer diffraction pattern produced by the aperture amplitude f at a distance z_0 from the aperture. f can be taken to describe the 'shape' of the aperture.

Apart from the factors in front of the integral, U is simply equal to the 2-D Fourier transform of f evaluated at the coordinates (noting that $k = 2\pi/\lambda$)

$$u = \frac{x_0}{z_0\lambda} \quad \text{and} \quad v = \frac{y_0}{z_0\lambda}$$

In practice, the observed quantity (optical field) is the intensity I given by $| U |^2$. We can therefore write

$$I(x_0, y_0) = \frac{1}{\lambda^2 z_0^2} | \tilde{\mathcal{F}}[f(x, y)] |^2$$

where $\tilde{\mathcal{F}}$ is the Fourier transform operator given by

$$\tilde{\mathcal{F}}[f(x, y)] = \int_{-\infty}^{\infty} f(x, y) \exp[-2\pi i(ux + vy)] dx dy$$

This result is the Fraunhofer diffraction formula and is based on the following assumptions:

(i) The aperture is thin (compared with the wavelength of light).

(ii) The diffraction pattern is observed in the far field.

Well known examples of Fraunhofer diffraction

Rectangular aperture $a \times b$

In this case,

$$f(x, y) = \begin{cases} 1, & | x | \leq \frac{a}{2}; \quad | y | \leq \frac{b}{2} \\ 0, & \text{otherwise} \end{cases}$$

and

$$\tilde{\mathcal{F}}[f(x, y)] = \int_{-a/2}^{a/2} \int_{-b/2}^{b/2} \exp[-i2\pi(ux + vy)] dx dy$$

$$= ab \operatorname{sinc}(\pi u a) \operatorname{sinc}(\pi v b)$$

The intensity pattern is therefore given by

$$I(x_0, y_0) = \frac{a^2 b^2}{\lambda^2 z_0^2} \operatorname{sinc}^2 \left(\frac{\pi a x_0}{\lambda z_0} \right) \operatorname{sinc}^2 \left(\frac{\pi a y_0}{\lambda z_0} \right)$$

Circular aperture with a radius of a

$$f(x,y) = \begin{cases} 1, & r \leq a \\ 0, & \text{otherwise} \end{cases} \qquad r = \sqrt{x^2 + y^2}$$

This involves the Hankel transform and a result for the intensity given by

$$I(r_0) = \frac{\pi^2 a^4}{\lambda^2 z_0^2} \left(\frac{2J_1(z)}{z} \right)^2$$

where $z = 2\pi a r_0 / \lambda z_0$ and $r_0 = \sqrt{x_0 + y_0}$. The first minimum occurs when $z = 3.83$, i.e. when

$$r_{\min} = \frac{1.22 \lambda z_0}{a}$$

This diffraction pattern is also the distribution seen at the focus of a well-corrected lens when a point monochromatic source is imaged. For a distant source, $z_0 \sim f$ - the focal length. Defining the F-number as $F \equiv f/a$, we get

$$r_{\min} = 1.22 \lambda F$$

The value of $r_{\min}$ determine the resolution of the image in the focal plane of the lens. The smaller the value of $r_{\min}$ the better the resolution. Thus, the smaller the F-number of a lens the greater its resolving power.

Fresnel Diffraction

Consider the Kirchhoff diffraction integral derived earlier

$$U(\mathbf{r_0}) = \frac{ik\alpha}{4\pi r_0} \iint_S \exp(i\mathbf{k} \cdot \mathbf{r}) \exp(ik \mid \mathbf{r} - \mathbf{r_0} \mid) dS$$

where

$$\alpha = \hat{\mathbf{n}} \cdot \hat{\mathbf{k}} + \hat{\mathbf{n}} \cdot \hat{\mathbf{r}}_0$$

Fresnel diffraction stems from considering the expansion of $|\mathbf{r}-\mathbf{r_0}|$ to second order and retaining the term $r^2/2r_0$, i.e.

$$\begin{aligned} |\mathbf{r}-\mathbf{r_0}| &= r_0 - \mathbf{r}\cdot\hat{\mathbf{r}}_0 + \frac{r^2}{2r_0} + \cdots \\ &\simeq r_0 - \mathbf{r}\cdot\hat{\mathbf{r}}_0 + \frac{r^2}{2r_0} \end{aligned}$$

This approximation is necessary when the diffraction pattern is observed in what is called the intermediate field or Fresnel zone.

$$U(\mathbf{r_0}) \simeq \frac{ik\alpha}{4\pi r_0}\exp(ikr_0)\iint_S \exp(i\mathbf{k}\cdot\mathbf{r})\exp(-ik\mathbf{r_0}\cdot\mathbf{r})\exp\left(ik\frac{r^2}{2r_0}\right)dS$$

Consider a Cartesian coordinate system

$$\mathbf{r} = \hat{\mathbf{x}}x + \hat{\mathbf{y}}y + \hat{\mathbf{z}}z$$

$$\mathbf{r_0} = \hat{\mathbf{x}}x_0 + \hat{\mathbf{y}}y_0 + \hat{\mathbf{z}}z_0$$

where x_0 and y_0 are taken to represent a position on a flat screen at a fixed distance z_0 from the aperture and let

$\hat{\mathbf{k}} \simeq \hat{\mathbf{z}}$	plane wave at normal incidence to the aperture
$\hat{\mathbf{r}}_0 \simeq \hat{\mathbf{z}}$	observations at small angles only
$kz \to 0$	'infinitely thin' aperture

Then,

$$U(x_0,y_0) = \frac{i}{\lambda}\frac{\exp(ikr_0)}{4\pi r_0}\iint_S \exp\left[-\frac{ik}{r_0}(xx_0+yy_0)\right]\exp\left[\frac{ik}{2r_0}(x^2+y^2)\right]dxdy$$

As with the analysis associated with Fraunhofer diffraction, we substitute

$$r_0 \simeq z_0 + \frac{x_0^2+y_0^2}{2z_0}$$

into the exponent $\exp(ikr_0)$ but use $r_0 \simeq z_0$ elsewhere. By introducing an aperture amplitude function $f(x,y)$, the amplitude function is then given by

$$U(x_0,y_0) = \frac{i}{\lambda}\frac{\exp(ikz_0)}{z_0}\exp\left(ik\frac{x_0^2+y_0^2}{2z_0}\right)$$

$$\times \int_{-\infty}^{\infty} f(x,y)\exp\left[-\frac{ik}{z_0}(xx_0+yy_0)\right]\exp\left[\frac{ik}{2z_0}(x^2+y^2)\right]dxdy$$

This result describes Fresnel diffraction - the diffraction pattern observed in the intermediate field. Apart from the factors in front of this integral, U is simply equal to the 2-D Fourier transform of the function

$$f(x,y)\exp\left[\frac{ik}{2z_0}(x^2+y^2)\right]$$

The term

$$\exp\left[\frac{ik}{2z_0}(x^2+y^2)\right]$$

is a quadratic approximation to a spherical wave. In simple terms

$$U \sim \text{spherical wave} \times \tilde{\mathcal{F}}[f \times \text{spherical wave}]$$

Noting that

$$\frac{ik}{2z_0}(x_0^2+y_0^2) + \frac{ik}{z_0}(-xx_0-yy_0) + \frac{ik}{2z_0}(x^2+y^2)$$

$$= \frac{ik}{2z_0}[x_0^2 - 2xx_0 + x^2 + y_0^2 - 2yy_0 + y^2]$$

$$= \frac{ik}{2z_0}[(x_0-x)^2 + (y_0-y)^2]$$

we can write the Fresnel diffraction formula in the form

$$U(x_0,y_0) = \frac{i}{\lambda}\frac{\exp(ikz_0)}{z_0}\int_{-\infty}^{\infty} f(x,y)\exp\left(\frac{i\pi}{\lambda z_0}[(x_0-x)^2+(y_0-y)^2]\right)dxdy$$

From the last equation we see that U is essentially (ignoring scaling constants) given by the convolution of the function

$$f(x,y)$$

with

$$\exp\left(\frac{ik}{2z_0}[x^2+y^2]\right)$$

or

$$U(x,y)=\frac{i}{\lambda}\frac{\exp(ikz_0)}{z_0}f(x,y)\otimes\otimes\exp\left(\frac{ik}{2z_0}[x^2+y^2]\right)$$

where $\otimes\otimes$ denotes the 2-D convolution integral.

Examples of Fresnel Diffraction

As an example, consider using the last expression for U to evaluate the Fresnel diffraction pattern from a square aperture of width a, i.e.

$$f(x,y)=\begin{cases}1, & |x|\leq\frac{a}{2},\ |y|\leq\frac{a}{2};\\ 0, & \text{otherwise.}\end{cases}$$

In this case, U is given by

$$U(x_0,y_0)=\frac{i}{\lambda}\frac{\exp(ikz_0)}{z_0}\int_{-\frac{a}{2}}^{\frac{a}{2}}\int_{-\frac{a}{2}}^{\frac{a}{2}}\exp\left(\frac{i\pi}{\lambda z_0}[(x_0-x)^2+(y_0-y)^2]\right)dxdy$$

which can be written in the form of two integrals I_1 and I_2; thus,

$$U(x_0,y_0)=\frac{i}{\lambda}\frac{\exp(ikz_0)}{z_0}I_1(x_0)I_2(y_0)$$

Since each integral is identical in form, we shall consider only I_1 given by

$$I_1(x_0)=\int_{-\frac{a}{2}}^{\frac{a}{2}}\exp\left[\frac{i\pi}{\lambda z_0}(x_0-x)^2\right]dx$$

Let

$$\xi = \sqrt{\frac{2}{\lambda z_0}}(x_0 - x)$$

so that

$$I_1(x_0) = \int_{\xi_1}^{\xi_2} \exp\left(i\frac{\pi}{2}\xi^2\right) d\xi \sqrt{\frac{\lambda z_0}{2}}$$

where

$$\xi_1 = \sqrt{\frac{2}{\lambda z_0}}\left(x + \frac{a}{2}\right) \quad \text{and} \quad \xi_2 = \sqrt{\frac{2}{\lambda z_0}}\left(x - \frac{a}{2}\right)$$

Defining the Fresnel integrals

$$C(\alpha) = \int_0^{\alpha} \cos\left(\frac{\pi t^2}{2}\right) dt, \quad S(\alpha) = \int_0^{\alpha} \sin\left(\frac{\pi t^2}{2}\right) dt$$

and noting that

$$\int_{\xi_1}^{\xi_2} = \int_0^{\xi_2} - \int_0^{\xi_1}$$

we can write

$$I_1(x_0) = \sqrt{\frac{\lambda z_0}{2}}([C(\xi_2) - C(\xi_1)] + i[S(\xi_2) - S(\xi_1)])$$

The amplitude field U is given by

$$U(x, y) = i\frac{\exp(ikz_0)}{2}([C(\xi_2) - C(\xi_1)] + i[S(\xi_2) - S(\xi_2)])$$

$$\times([C(\eta_2) - C(\eta_1)] + i[S(\eta_2) - S(\eta_1)])$$

where

$$\eta_1 = \sqrt{\frac{2}{\lambda z_0}}\left(y + \frac{a}{2}\right) \quad \text{and} \quad \eta_2 = \sqrt{\frac{2}{\lambda z_0}}\left(y - \frac{a}{2}\right)$$

The corresponding intensity is

$$I(x_0, y_0) = \frac{1}{4}(|\, L_\xi \,|^2 |\, L_\eta \,|^2)$$

where $L(\xi)$ denotes the length of the vector

$$[C(\xi_2) - C(\xi_1)] + i[S(\xi_2) - S(\xi_1)]$$

The characteristic behaviour of this vector is represented graphically by the Cornu spiral. The Cornu spiral is a plot of $S(\alpha)$ vs. $C(\alpha)$ and can be used to graphically solve for the complex amplitude or intensity in a Fresnel diffraction problem.

On the Fourier Transforming Properties of a Lens

At the plane of focus of a well-corrected lens, the following complex amplitude is observed (Fraunhofer condition):

$$U(x_0, y_0) = \frac{i}{\lambda} \frac{\exp(ikz_0)}{z_0} \exp\left(ik \frac{x_0^2 + y_0^2}{2z_0}\right) \tilde{\mathcal{F}}[f(x,y)]$$

Thus, if a transparency whose amplitude transmittance is $t(x,y)$ is placed just in front of a lens and then illuminated by a unit plane wave at normal incidence, the observed complex amplitude at the focal plane is given by

$$U(x_0, y_0) = \frac{i}{\lambda} \frac{\exp(ikf)}{f} \exp\left(ik \frac{x_0^2 + y_0^2}{2f}\right) \tilde{\mathcal{F}}[t(x,y)]$$

where f is the focal length of the lens and

$$\tilde{\mathcal{F}}[t(x,y)] = \int_{-\infty}^{\infty} t(x,y) \exp[2\pi i(ux + vy)] dx dy; \quad u = \frac{x_0}{\lambda f}, \quad v = \frac{y_0}{\lambda f}$$

The amplitude is therefore almost given by a Fourier transform - there is an additional quadratic phase factor - and the intensity is given by

$$I(x,y) = \frac{1}{\lambda^2 f^2} \mid \tilde{\mathcal{F}}[t(x,y)] \mid^2$$

It is interesting to ask whether a transparency can be placed a distance d in front of the lens so as to give an exact Fourier transform? The amplitude $T(x,y)$ generated by the transparency in front of the lens is given by the Fresnel diffraction formula

$$T(x,y) = \frac{i}{\lambda}\frac{\exp(ikd)}{d} t(x,y) \otimes\otimes \exp\left[\frac{i\pi}{\lambda d}(x^2+y^2)\right]$$

and a lens performs a Fourier transform on the field T. Therefore in the focal plane

$$U(x_0,y_0) = \frac{i}{\lambda}\frac{\exp(ikf)}{f} \exp\left(ik\frac{x_0^2+y_0^2}{2f}\right) \tilde{\mathcal{F}}[T(x,y)]$$

Using the convolution theorem and noting that

$$\tilde{\mathcal{F}}\left(\exp\left[\frac{i\pi}{\lambda d}(x^2+y^2)\right]\right) = \lambda d \exp[-i\pi\lambda d(u^2+v^2)]$$

we have

$$\tilde{\mathcal{F}}[T(x,y)] = \tilde{T}(u,v) = i\frac{\exp(ikd)}{d}\tilde{t}(u,v)\exp[-\pi i\lambda d(u^2+v^2)]$$

where $\tilde{t} = \tilde{\mathcal{F}}[t]$ and

$$\exp[-\pi i\lambda d(u^2+v^2)] = \tilde{\mathcal{F}}\left(\exp\left[\frac{i\pi}{\lambda d}(x^2+y^2)\right]\right)$$

The field U is therefore given by

$$\begin{aligned} U(x_0,y_0) &= -\frac{1}{\lambda f}\exp[ik(f+d)]\exp\left[\frac{i\pi}{f\lambda}(x_0^2+y_0^2)\right] \\ &\quad \times \exp\left[-i\pi\frac{\lambda d}{(\lambda f)^2}(x_0^2+y_0^2)\right]\tilde{t}(u,v) \\ &= -\frac{1}{\lambda f}\exp[ik(f+d)]\exp\left[\frac{\pi i}{\lambda f}(x_0^2+y_0^2)\left(1-\frac{d}{f}\right)\right]\tilde{t}(u,v) \end{aligned}$$

Now, when $d = f$

$$U(x_0,y_0) = -\frac{1}{\lambda f}\exp[ik(f+d)]\tilde{t}(u,v)$$

can be used to filter the spatial frequency components of a source distribution via the application of certain optical masks in the focal plane of the first lens - spatial filtering. For example, spatial filtering can restore the quality of a collimated laser beam by blocking all spatial frequencies due to the interaction of the beam with dust particles. In general, any 'noise' induced by the (multiple) scattering of light with a complex of subwavelength objects (such as dust particles) modifies the spatial frequencies with high values and so can be removed by application of an optical lowpass filter (a mask placed at the focal plane of a lens).

In general, the use of a lens system to generate the Fourier transform of an image (forward Fourier transform) and to recover an image from this transform (inverse Fourier transform) provides an optical method of processing signals and images. This is known as optical signal processing. Optical signal processing exploits the Fourier transforming properties of a lens to process (optically filter) an image in the same way that a fast Fourier transform can be used to process (digitally filter) a digital signal.

Optical Systems and Image Formation

An optical 'system' may be defined as that which produces a set of output functions from a set of input functions. Physically, it may be an electrical circuit (inputs and outputs are voltages) or an optical system where the inputs and outputs are either complex amplitudes or intensities. From the point of view of 'linear systems theory', the physical nature of the system is unimportant. A general system can be represented by the operator $\hat{S}$:

$$g(x,y) = \hat{S}[f(x,y)]$$

A linear system has the property that

$$\hat{S}[af_1(x,y)+bf_2(x,y)] = a\hat{S}[f_1(x,y)]+b\hat{S}[f_2(x,y)]$$

for all inputs f_1 and f_2 and all constants a and b. Linearity implies that an output function can be broken down into elementary functions, each of which can be separately passed through the system; the total output is then the sum of the 'elementary' outputs.

The 'shifting property' of the delta function allows us to consider any input function to be a linear combination of weighted and displaced delta functions:

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x',y')\delta(x-x')\delta(y-y')dx'dy'$$

giving an output

$$g(x,y) = \hat{S}[f(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x',y')\hat{S}[\delta(x-x')\delta(y-y')]dx'dy'$$

The system response at (x,y) due to a delta function input at (x',y') is called the impulse response function

$$p(x,y;x',y') = \hat{S}[\delta(x-x')\delta(y-y')]$$

In optical as well as many other imaging systems, the quantity p is called the point spread function. For a linear optical system:

$$g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x',y')p(x,y;x',y')dx'dy'$$

Note that the optical diffraction formulae are of the same form where for Fraunhofer diffraction

$$p(x,y;x',y') = \frac{i}{\lambda}\frac{\exp(ikz)}{z}\exp\left(ik\frac{(x^2+y^2)}{2z}\right)\exp\left[-\frac{ik}{z}(xx'+yy')\right]$$

and for Fresnel diffraction

$$p(x,y;x',y') = \frac{i}{\lambda}\frac{\exp(ikz)}{z}\exp\left(\frac{i\pi}{2z}[(x-x')^2+(y-y')^2]\right)$$

If the impulse response function of a linear system depends only on the coordinate differences $(x-x')$ and $(y-y')$, and not on each coordinate separately, i.e.

$$p(x,y;x',y') \equiv p(x-x',y-y')$$

then we obtain an expression for g which involves the simple convolution relationship

$$g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x',y')p(x-x',y-y')dx'dy'$$

Such a system is called stationary. In optical imaging, a stationary optical system is called isoplanatic. Isoplanicity requires that the point spread function is the same for all field angles and implies that the aberrations are independent of field angle. Many real imaging systems are (to a good approximation) both linear and isoplanatic.

The convolution relationship between input and output suggests using Fourier transform (FT) techniques which, via the convolution theorem, give

$$G(u,v) = F(u,v)T(u,v)$$

where

$$T(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\exp[-2\pi i(ux+vy)]dxdy$$

i.e.

FT of output = (FT of input) × (FT of impulse response function)

The quantity T is called the system transfer function. In optical imaging systems, T is called the optical transfer function or OTF. The OTF is just the 2-D FT of the point spread function. Note that

(i) The convolution relationship only applies to linear stationary optical systems.

(ii) There is no unique OTF for an optical system with field-dependent aberrations (non-stationary).

(iii) There is no unique OTF for an optical system when an object is illuminated by spatially partially coherent light (non-linear system in both complex amplitude and intensity).

Incoherent Image Formation

Consider the case where the object plane is illuminated by a plane or spherical wave - by perfectly spatially coherent light. Let the complex amplitude immediately after the object be denoted by $U_{\text{in}}(x, y)$ and $U_{\text{out}}(x, y)$ be the complex amplitude at the image plane and let the complex amplitude at (x, y) in the output due to a unit strength point at in the input be $p(x, y; x', y')$. The total amplitude at (x, y) due to all such points in the object plane is then given by

$$U_{\text{out}}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_{\text{in}}(x', y') p(x, y; x', y') dx' dy'$$

For an isoplanatic optical system, this reduces to

$$U_{\text{out}}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_{\text{in}}(x', y') p(x - x', y - y') dx' dy'$$

Note that a spatially coherent optical system is linear in the complex amplitude. Consider the case of narrowband light that is not perfectly spatially coherent. The general complex representation of the time-varying scalar field

is called the analytic signal $V(\mathbf{r},t)$; it is defined such that

$$\text{realscalarfield} = \Re[V(\mathbf{r},t)]$$

For narrowband light, the analytic signal can be written as a product of a 'slowly varying' function - the time-varying complex amplitude $U(\mathbf{r},t)$ times $\exp(-i\omega t)$, i.e.

$$V(\mathbf{r},t) = U(\mathbf{r},t)\exp(-i\omega t)$$

The instantaneous intensity is defined as

$$I(\mathbf{r},t) = \mid U(\mathbf{r},t) \mid^2$$

whereas the time-averaged intensity $\bar{I}(\mathbf{r})$ (i.e. that observed by an optical detector), is equal to

$$\bar{I}(\mathbf{r}) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} I(\mathbf{r},t)dt$$

In general, the time-varying complex amplitudes are related by

$$U_{\text{out}}(x,y,t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U_{\text{in}}(x',y',t)p(x,y;x'y')dx'dy'$$

Coherent illumination implies $U(x,y,t) = U(x,y)$ does not vary in time. The average intensity is therefore given by

$$\begin{aligned}
\bar{I}_{\text{out}}(x,y) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mid U_{\text{out}}(x,y,t) \mid^2 dt \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y;x',y')p^*(x,y;x',y') \\
&\quad \times \left[\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} U_{\text{in}}(x',y',t)U_{\text{in}}^*(x'',y'',t)dt\right] dx'dy'dx''dy''
\end{aligned}$$

The term in [ ] is called the mutual intensity of the narrowband light

$$J_{\text{in}}(x', y'; x'', y'') = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} U_{\text{in}}(x', y', t) U_{\text{in}}^*(x'', y'', t) dt$$

or

$$J_{\text{in}}(\mathbf{r}', \mathbf{r}'') = \langle U_{\text{in}}(\mathbf{r}', t) U_{\text{in}}^*(\mathbf{r}'', t) \rangle$$

Incoherent light is defined to be such that

$$J(\mathbf{r}', \mathbf{r}'') = \bar{I}(\mathbf{r}') \delta(\mathbf{r}' - \mathbf{r}'')$$

In other words, two neighbouring points $\mathbf{r}'$ and $\mathbf{r}''$ have uncorrelated fields, for any $\mathbf{r}' \neq \mathbf{r}''$. Using the definition for incoherent light above, the expression for $\bar{I}_{\text{out}}$ becomes

$$\begin{aligned} \bar{I}_{\text{out}}(x, y) \;=\; & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y; x', y') p^*(x, y; x', y') \\ & \times \bar{I}_{\text{in}}(x', y') \delta(x' - x'') \delta(y' - y'') dx' dy' dx'' dy'' \end{aligned}$$

or

$$\bar{I}_{\text{out}}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} | \, p(x, y; x', y') \, |^2 \, \bar{I}_{\text{in}}(x', y') dx' dy'$$

The quantity $|p(x, y; x', y')|^2$ is the intensity point spread function. For an isoplanatic optical system, this result reduces to

$$I_{\text{out}}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{I}_{\text{in}}(x', y') \, | \, p(x - x', y - y') \, |^2 \, dx' dy'$$

where the bar over I has been omitted when referring to the intensity - a time average is always assumed.

For perfectly incoherent illumination, an optical system is linear in intensity and, if isoplanicity holds, the output (image) intensity is equal to the input (object) intensity convolved with the intensity point spread function.

If p is normalized to unit volume (it is, since $P(0,0) = 1$), $| p |^2$ is not, so we normalize it by dividing by its infinite integral:

$$\frac{| p(x,y) |^2}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} | p(x,y) |^2 \, dxdy}$$

The incoherent optical transfer function (IOTF) $T(u,v)$ is the Fourier transform of the (normalized) point spread function; applying the autocorrelation theorem to the top line and Parseval's theorem to the bottom line,

$$T(u,v) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(\xi,\eta)P^*(\xi+\lambda zu, \eta+\lambda zv)d\xi d\eta}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} | P(\xi,\eta) |^2 \, d\xi d\eta}$$

where $P(\xi,\eta)$ is the inverse Fourier transform of $p(x,y)$,

$$P(\xi,\eta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y) \exp\left[i\frac{2\pi}{\lambda z}(\xi x + \eta y)\right] dxdy$$

and is the pupil function. The equation above for T basically states that the IOTF is equal to the (normalized) spatial autocorrelation of the pupil function. Note, that the IOTF relates the input and output intensity spectra,

$$\tilde{I}_{\text{out}}(u,v) = \tilde{I}_{\text{in}}(u,v)T(u,v)$$

The spatial frequencies are intensity frequencies and are not the same as the amplitude frequencies produced in a coherent optical system. This expression for T can be written in the form

$$T(u,v) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P\left(\xi-(\lambda zu)/2, \eta-(\lambda zv)/2\right) P^*\left(\xi+(\lambda zu)/2, \eta+(\lambda zv)/2\right) d\xi d\eta}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} | P(\xi,\eta) |^2 \, d\xi d\eta}$$

From this result it follows that:

(i) $T(0,0) = 1$ (because of normalization).

(ii) $T(-u,-v) = T^*(u,v)$ (Fourier transform of a real quantity).

(iii) $T(u,v) \leq T(0,0)$ (using the Schwartz inequality).

The modulus $\mid T \mid$, or modulation transfer function (MTF), describes the transfer or modulation of sinusoidal components of the object. The phase of T, describes spatial shifts of the sinusoidal components.

Coherent Image Formation

With coherent light, the complex amplitude of the image is equal to that at the object plane convolved with the amplitude point spread function (for an isoplanatic system):

$$U_{\text{out}}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U_{\text{in}}(x',y')h(x-x',y-y')dx'dy'$$

where

$$p(x,y) = C\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(u,v)\exp\left[-i\frac{2\pi}{\lambda z}(ux+vy)\right]dudv$$

and P is the pupil function of the optical system, i.e. the complex amplitude in the exit pupil. The constant C is usually chosen so that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)dxdy = P(0,0) = 1$$

The pupil function P is, for a clear pupil, defined by

$$P(u,v) = \begin{cases} \exp[ikW(u,v)], & (u,v) \in \text{aperture} \\ 0, & \text{otherwise} \end{cases}$$

so that $P(0,0) = 1$ and therefore $C = 1$. The function W is called the wave aberration function. A shaded or apodized pupil can be handled by introducing an absorption term A thus:

$$P(u,v) = A(u,v)\exp[ikW(u,v)]$$

Taking the Fourier transform of U_{out} and using the convolution theorem we can write

$$\widetilde{U_{\text{out}}}(u,v) = \widetilde{U_{\text{in}}}(u,v)T(u,v)$$

where $\widetilde{U_{\text{out}}}$ is the spectrum of image amplitude, $\widetilde{U_{\text{in}}}$ is the spectrum of object amplitude and T is the coherent optical transfer function (COTF). Note that

$$T(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\exp[-2\pi i(ux+vy)]dxdy = P(\lambda zu, \lambda zv)$$

i.e. the COTF at spatial frequency (u,v) is simply equal to the pupil function at coordinates $(\lambda zu, \lambda zv)$.

The intensity distribution $I(x,y)$ of a coherent image is given by,

$$\begin{aligned} I(x,y) &= |U_{\text{out}}|^2 \\ &= |p(x,y) \otimes\otimes U_{\text{in}}(x,y)|^2 \end{aligned}$$

The model can be used to develop computer generated speckle patterns where $U_{\text{in}}(x,y)$ is constructed using white Gaussian noise. Physically the noise function represents a first approximation to a (rough) surface from which coherent light is scattered.

Appendix C

Simulation of the Lippmann Interferential Photographic Technique (Paper)

Simulation of the Lippmann Interferential Photographic Technique

Hans I. Bjelkhagen and Savita De Souza
Modern Optics, SERC, Hawthorn Building, De Montfort University, Leicester LE1 9BH

Abstract

Lippmann photography is an old colour photographic technique invented in 1891 by Gabriel Lippmann. Lippmann developed the first theory of recording monochromatic and polychromatic spectra. The current paper describes a new optical security application of this technique as well as describes the theory behind the simulation of the Lippmann process. Currently, this type of technique can be applied as a unique security device on security documents, like passports, identification cards, credit cards, driving licenses etc. This method offers a very high degree of security and has many advantages. Lippmann photographs are very similar to holograms currently used in this field, but this technique gives a unique recording of each document so that a much higher degree of security can be achieved.

1 Introduction

Lippmann photography is an old colour photographic technique invented in 1891 by Gabriel Lippmann. Lippmann developed the first theory of recording monochromatic and polychromatic spectra. This technique is mainly the photographic recording of an interference or standing wave-pattern caused by interference between incoming light and its own reflection. Lippmann interferential photography was the first direct way to make colour photographs. The current project is based on a new optical security application of this technique. Currently, this type of technique can be applied as a unique security device on security documents, like passports, identification cards, credit cards, driving licences etc. This method offers a very high degree of security and has many advantages. To list a few of these advantages:

- The recording of a Lippmann OVD (Optical Variable Device) is rather simple to perform, as no specially equipped laboratory is required.
- The cost of a Lippmann OVD is low.
- The Lippmann OVD has a very high archival stability.
- The Lippmann OVD is Bragg sensitive, which means it changes its colour depending on the angle of illumination and observation.
- The Lippmann OVD can record latent images, used as security devices on documents.

- The Lippmann OVD cannot be copied by conventional colour photography nor can it be copied on colour copy machines.
- The manufacturer of the film can strictly control the access to the special photopolymer material required for the recording.
- The Lippmann OVD can be laminated to a light absorbing material, for example a black plastic foil, which means that it is not possible to see through the Lippmann film.
- Since the resolution of the Lippmann OVD is extremely high, a reduced image of the security document can be laminated to the document, occupying only a limited area of it.

Lippmann photographs are very similar to holograms currently used in this field, but this technique gives a unique recording of each document so that a much higher degree of security can be achieved. It is based on Bragg diffraction from photographically recorded volume gratings, in which the index and absorption variations occur mainly in the direction normal to the film plane. The Lippmann OVD is a new and unique type of OVD, which belongs to the interference security image structures (ISISs). The main advantage of this device is that it can be individually made, a unique label for each document.

2 Brief description of the photopolymer recording technique

A Lippmann photograph can be recorded on a photopolymer material in the following way. The photosensitive polymer layer must be rather thin, of the order of a few micrometers only. The photopolymer layer must be coated on a flexible transparent base and a special type of reflecting foil must be laminated on top of the photosensitive polymer layer in absolutely perfect contact with it. A preferred choice is the colour photopolymer material with an emulsion thickness of about 2 to 3 μm. DuPont has manufactured such experimental materials. The polymer film laminated to the reflecting foil must be exposed in a special camera. If the recording material is not perfectly isochromatic, a correction filter may be required in front of the camera lens to obtain correct colour balance. After being exposed to the image-forming information in the camera, the film must be processed. The reflecting foil is detached from the photopolymer film and the photopolymer layer must be exposed to strong white light or UV light for development. After that, the photograph is put in an oven for a certain time to increase the brightness of the image. The whole processing technique of the photopolymer film is completely dry. Based on this fact, an automatic recording and processing machine for the recording of Lippmann security labels can be developed. After being processed, the transparent photopolymer label is laminated to its corresponding security document. The polymer film contains no dyes or any fading chemicals, which means that the archival stability is expected

to be very high. The photograph is simply a piece of plastic material with the information recorded in it as an optical phase structure, refractive index variations within the photopolymer layer.

3 Principle of Lippmann Photography

The recording material has a rather low light sensitivity due to the demand for high resolving power for recording Lippmann photographs. The photosensitive emulsion coated on Lippmann plates is brought in contact with a highly reflecting surface. Lippmann used mercury in contact with the emulsion. This mirror reflects the light into the emulsion, which then interferes with the light coming from the other side of the emulsion. Stationary standing waves of the interfering light produce a very fine fringe pattern throughout the emulsion with a periodic spacing of $\lambda/(2n)$ that is recorded (λ is the wavelength of light in air and n is the refractive index of the emulsion). The colour information concerning the object is recorded in this way. For example a large separation between the fringes in the emulsion indicates that the recorded wavelength is located at the red end of the spectrum. More closely spaced fringes indicate a shorter wavelength, such as green or blue. The description applies only when rather monochromatic colours are recorded. When the developed photograph is viewed in white light, different parts of the recorded image produce different colours due to the separation of the recorded fringes in the emulsion. The light is reflected from the fringes, creating different colours corresponding to the original ones that produced them during the recording. It is obvious that there is a severe requirement on the resolving power to record the fringes separated in the order of half the wavelength of the light. It is also clear that the processing of these plates is critically important, as one cannot change the separation between the fringes since that would create wrong colours. To observe the correct colours, the illumination and the observation must be at normal incidence. The colour of the image changes with the change in colour. The change in colour is known as iridescence, which is a very important feature of the Lippmann photograph and is an important attribute of the Lippmann photograph as a document security device. Thus the Lippmann image is recorded as a Bragg structure in the film. Lippmann photography involves no phase recording; the recorded interference structure is a result of phase-locking the light by the reflecting mirror. Figure 1 shows the principle of the technique.

A special recording-processing device must be designed and manufactured in which a roll of the photosensitive film, prelaminated with a reflecting foil, is exposed. Special illuminating lamps, such as strong halogen lamps are required to record the security documents with a reasonably short exposure time. There are two modern materials suitable for recording, panchromatic ultra-high-resolution silver-halide materials and panchromatic photopolymer materials. After being exposed to the image-forming information in the camera, the Lippmann film must be processed. The processing technique of the photopolymer film is com-

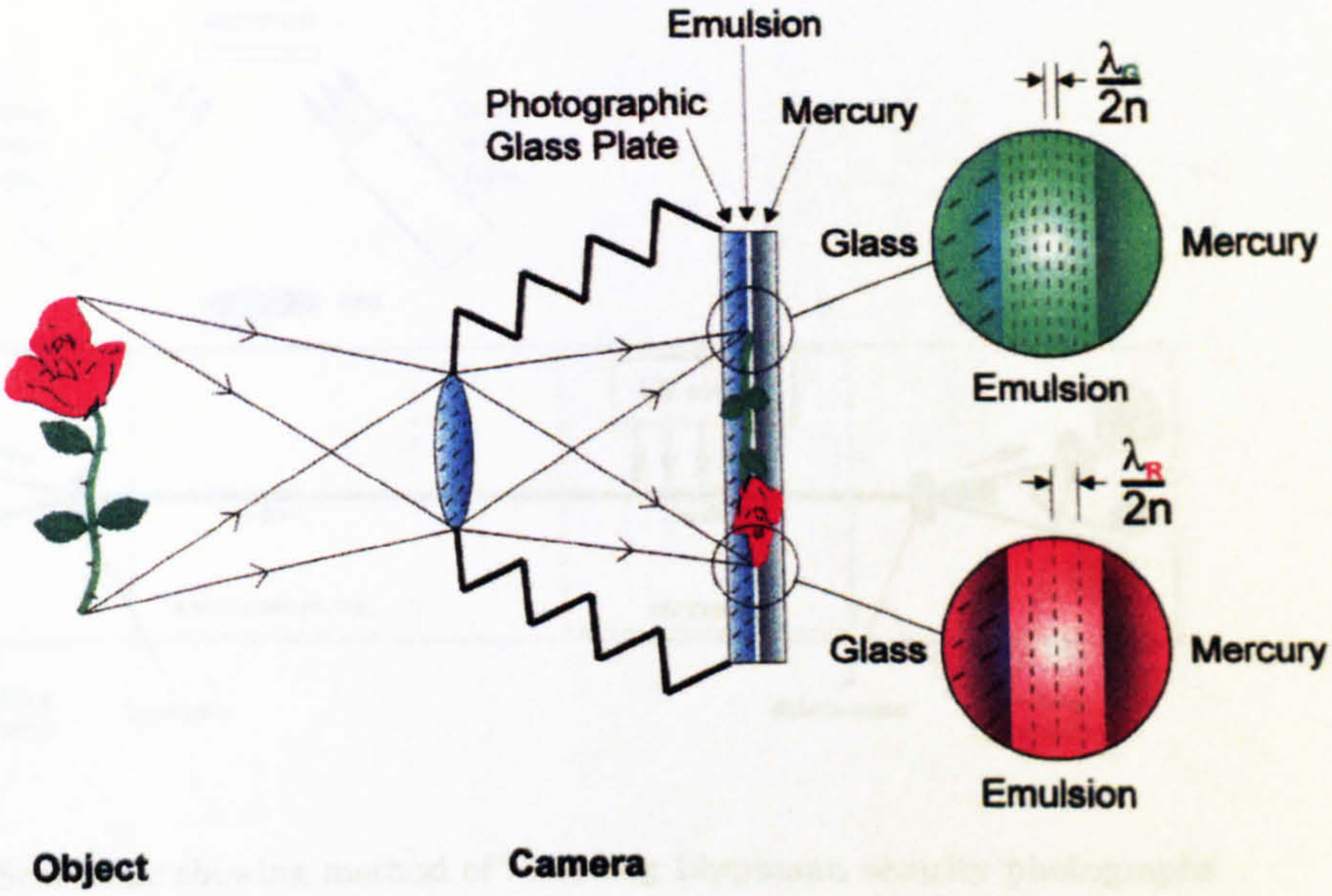


Figure 1. Principle of Lippmann Photography

pletely dry. Hence due to this fact, an automatic recording and processing machine for the production of Lippmann security labels can be developed. After being processed, the transparent photopolymer label is laminated to its corresponding security document. The polymer film contains no dyes or any fading chemicals, which means that the archival stability is expected to be very high. The photograph is basically a piece of plastic material with the information recorded in it as an optical phase structure. Figure 2 shows the method of recording Lippmann security photographs.

4 Security Application of the technique

The procedure is as follows: First the document information has to be exposed to the recording film, using a special camera. Then the film has to be processed. There are two methods for this depending on the material used, photopolymer materials are processed using dry processing and silver-halide materials use wet processing. An automatic processing machine is required. After the processing is complete, the recorded images can be laminated to a black backing plastic foil. Then the Lippmann image is laminated to the corresponding security documents. The figure 3 shows a Lippmann photograph attached to a security document. This figure is Bragg sensitive and hence perpendicular observation of the image gives the correct colour and oblique observation shifts the colour to-

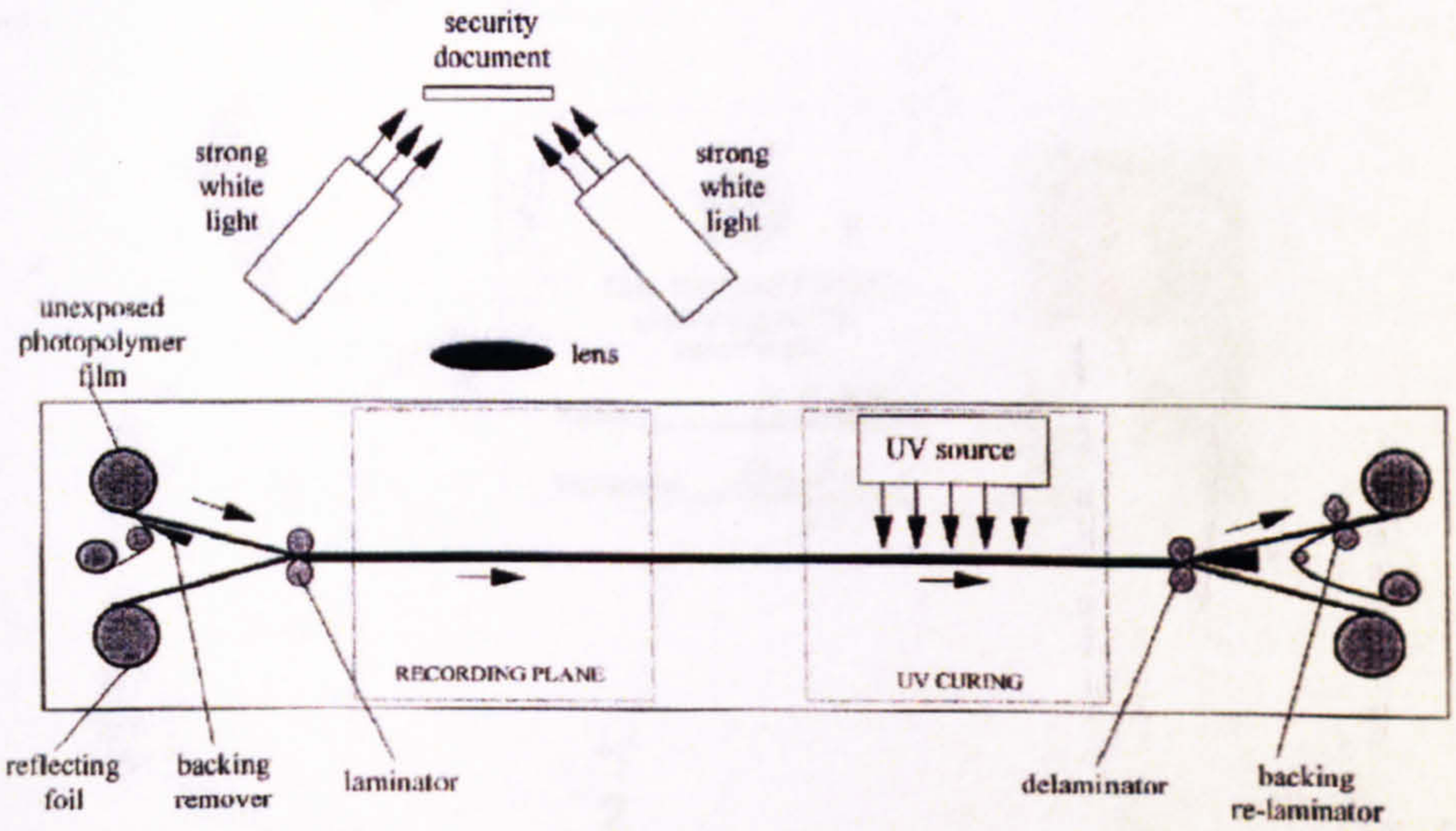


Figure 2. Schematic showing method of recording Lippmann security photographs

wards the shorter wavelength. This is a very important feature of the Lippmann photographs and hence it makes it virtually impossible to copy such images on conventional photocopiers or even to replace the Lippmann photograph with an ordinary photograph.

5 Simulation of the Lippmann process

It has been determined that the calculations for the simulation process are one-dimensional calculations.

The transmission line model is a very powerful method for analysing 1D wave propagation in stratified media. Mathematically it is equivalent to the transfer matrix method but instead of 2×2 matrices one uses recursion relations, which yield simpler algorithms, especially for reflection coefficient calculations.

6 Reflection coefficient by the transmission line model

We first show how the plane wave response of a vertically stratified medium can be computed by use of the transmission line model. We only consider plane waves in the $\pm z$ direction but the method is readily generalized to arbitrary incident fields by means of a Fourier expansion of the x, y dependence (the angular spectrum expansion). We assume the stratified medium to consist of N layers of finite thickness on top of an infinitely thick bottom layer. The m^{th} layer ($m = 0, 1, 2 \ldots N$) has thickness d_m ($d_N \to \infty$), a refractive index $n_m = \sqrt{\epsilon_m/\epsilon_0}$,

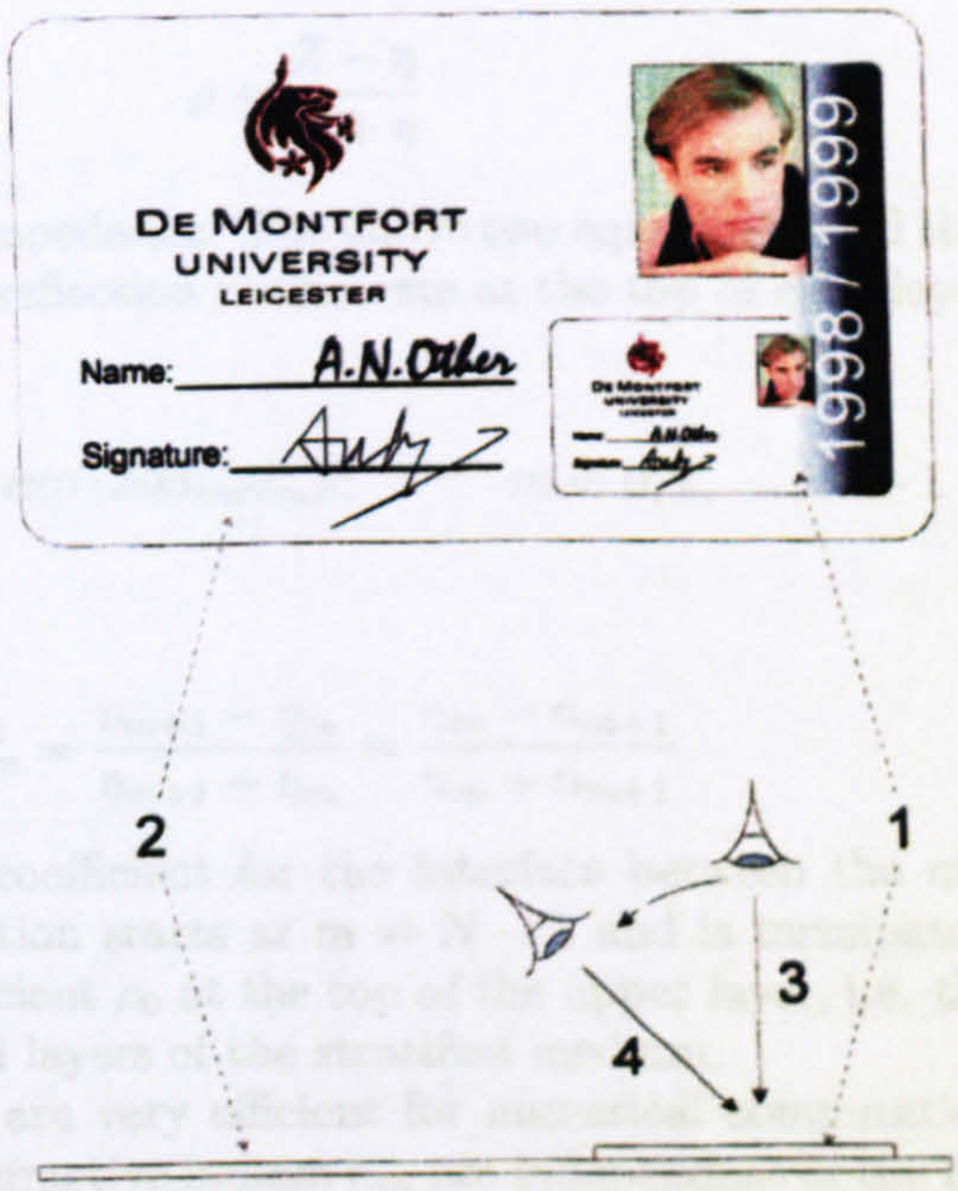


Figure 3. Lippmann photograph attached to a security document

and a wave impedance $\eta_m = \eta_0/n_m$, where $\eta_0 = \sqrt{\mu_0/\epsilon_0} \approx 120\pi\Omega$ is the free space wave impedance. At any point in the medium the impedance Z is defined as the ratio between the tangential components of the electric and the magnetic fields of the *composite* wave, i.e., the superposition of the up- and down-going plane waves. The impedance combines the two quantities that are continuous across the layer interfaces. With Z_m being the impedance at the top of the m^{th} layer, we have the following, well known iteration formula:

$$Z_m = \eta_m \frac{1 + \frac{Z_{m+1}-\eta_m}{Z_{m+1}+\eta_m} \exp(2i\delta_m)}{1 - \frac{Z_{m+1}-\eta_m}{Z_{m+1}+\eta_m} \exp(2i\delta_m)}; \qquad m = 0, 1, \ldots, N-1$$

where $\delta_m = kn_m d_m$ is the transmission "phase-shift" for the m^{th} layer (the wave number being $k = \omega\sqrt{\mu_0\epsilon_0} = \omega/c$). Since we only have a down-going wave in the infinite bottom layer, the impedance there simply equals the wave impedance of the layer: $Z_N = \eta_N$. Starting at the top of the bottom layer, we can use the above equation to compute the impedance at all the interfaces iteratively, until we finally reach that of the surface layer. For our purpose it is more convenient to work with the reflection coefficient. It is defined as the complex ratio between

the electric fields of the up- and down-going waves. At any point the reflection coefficient is related to the impedance Z by

$$\rho = \frac{Z - \eta}{Z + \eta}$$

where η is the local wave impedance. The above two equations yield the following recursion formula for the reflection coefficients at the top of each layer:

$$\rho_m = \frac{\rho_m^0 + \rho_{m+1}}{1 + \rho_m^0 \rho_{m+1}} \exp(2ikn_m d_m); \qquad m = 0, 1, \ldots, N-1 \tag{6.1}$$

where $\rho_N = 0$ and

$$\rho_m^0 = \frac{\eta_{m+1} - \eta_m}{\eta_{m+1} + \eta_m} = \frac{n_m - n_{m+1}}{n_m + n_{m+1}} \tag{6.2}$$

is the Fresnel reflection coefficient for the interface between the m^{th} and the $m+1^{\text{th}}$ layer. The iteration starts at $m = N-1$ and is terminated when we reach the reflection coefficient ρ_0 at the top of the upper layer, i.e. the resulting reflection coefficient of all layers of the stratified medium.

The above equations are very efficient for numerical computations. In the non-dispersive case the refractive indices n_m are independent of the frequency so the frequency dependence enters only through the wave number k. In a simple MathCAD implementation the reflection coefficients ρ_0 from 174 layers and for 1024 different frequencies are computed in about 7 sec.

In the low frequency limit $k \to 0$, the first equation yields $Z_m = Z_{m+1}$ which implies that $Z_0 = Z_N = \eta_N$ and:

$$\lim_{k \to 0} \rho_0 = \frac{\eta_N - \eta_0}{\eta_N + \eta_0} \tag{6.3}$$

This feature should be retained also in approximate models.

7 The first Born approximation

The equation (6.1) shows that ρ_0 is generally a non linear function of the parameters of the medium. However, a linear model is obtained in the weak scattering or first Born approximation. It applies for $|\rho_m^0|, |\rho_m| \ll 1$, in which case we can use the approximation

$$\frac{\rho_m^0 + \rho_{m+1}}{1 + \rho_m^0 \rho_{m+1}} \cong \rho_m^0 + \rho_{m+1}$$

in Equation 6.1. We then obtain

$$\rho_m \cong (\rho_m^0 + \rho_{m+1}) \exp(2ikn_m d_m)$$

and the iteration becomes trivial, yielding:

$$\rho_0^B = \sum_{m=0}^{N-1} \rho_m^0 \exp\left(2ik \sum_{s=0}^{m} n_s d_s\right)$$

where the superscript ^{B} is used to indicate the first Born approximation. In the low frequency limit $k \to 0$, we now obtain

$$\rho_0^B \to \sum_{m=0}^{N-1} \rho_m^0 \neq \frac{\eta_N - \eta_0}{\eta_N + \eta_0} \tag{7.1}$$

which, however, is not consistent with Equation(6.3). This inconsistency is removed when the Fresnel coefficients in equation 6.2 are approximated by:

$$\rho_m^0 = \frac{\eta_{m+1} - \eta_m}{\eta_{m+1} + \eta_m} \cong \frac{1}{2} \ln\left(\frac{\eta_{m+1}}{\eta_m}\right)$$

Then equation 7.1 yields

$$\rho_0^B \to \sum_{m=0}^{N-1} \rho_m^0 = \frac{1}{2} \ln\left(\frac{\eta_N}{\eta_0}\right) \cong \frac{\eta_N - \eta_0}{\eta_N + \eta_0}$$

The consistent first Born approximation is therefore given by:

$$\rho_0^B = \sum_{m=0}^{N-1} \frac{1}{2} \ln\left(\frac{\eta_{m+1}}{\eta_m}\right) \exp\left(2ik \sum_{s=0}^{m} n_s d_s\right)$$

$$= -\sum_{m=0}^{N-1} \frac{1}{2} \ln\left(\frac{n_{m+1}}{n_m}\right) \exp\left(2ik \sum_{s=0}^{m} n_s d_s\right)$$

In the continuous limit ($d_m \to dx, N \to \infty$) we now obtain the usual formula:

$$\rho_0^B(k) = -\frac{1}{2} \int_0^\infty dz \frac{d \log n(z)}{dz} \exp\left(2ik \int_0^z n(z')dz'\right)$$

At this point we introduce the two-way optical path-length:

$$u = u(z) = 2 \int_0^z n(z')dz'$$

as a new integration variable. We then obtain:

$$\rho_0^B(k) = -\frac{1}{2} \int_0^\infty du \frac{d \ln n(u)}{du} \exp(iku)$$

$$= \frac{1}{2} ik \int_0^\infty du \ln[n(u)] \exp(iku)$$

where the last expression follows from an integration by parts and the fact that we have assumed $n(0) = 1$.

8 Display Unit for the Lippmann Photographs

A display device based on a Visual Plus Slide Illuminator turned upside down, mounted in a holder, has been made. It is used for convenient viewing of the security documents. This provides uniform white light and will enable, e.g., the passport officer to check the documents with ease to detect any fraud. Figure 4 demonstrates the use of the display unit.

Figure 4. Display Unit for the Lippmann Photographs

Bibliography

1. H.I. Bjelkhagen, 'The Lippmann OVD for enhanced document security', *Proc. SPIE* **3973**, 276-283 (2000).

2. H.I. Bjelkhagen, 'New optical security device based on one-hundred-year-old photographic technique', *Opt. Eng.* **38**, 55-61 (1999).

3. Herbert E. Ives, 'An experimental study of the Lippmann Colour Photograph', *Astrophysical Journal* **27**, 325-352 (1908).

4. H.I. Bjelkhagen, 'Lippmann Photography, Reviving an early colour process', *History of Photography* **23** (3), (1999).

5. P. Connes, 'Silver salts and standing waves: The History of Interference Colour Photography', *J. Optics.* **18**, 147-166 (1987).

6. Jean-Marc Fournier, Paul L. Burnett, 'Color Rendition and Archival Properties of Lippmann Photographs', *Journal of Imaging Science and Technology* **38**, (6) (1994).

7. Helge Nareid, Hans M. Pedersen, 'Modelling of the Lippmann Colour Process', *J. Opt. Soc. Am. A.* **8**, (2) (1991).

8. P. Hariharan, 'Lippmann photography or Lippmann holography?', *Journal of Modern Optics* **45**, (8) 1759-1762 (1998).

9. N.J. Phillips, R.A.J. Van Der Werf, 'The Creation of Efficient Reflective Lippmann Layers in Ultra-Fine Grain Silver Halide Materials using Non-Laser Sources', *The Journal of Photographic Science* **33**, (1985).

10. R.E. Schwall, P.D. Zimmerman, 'Lippmann Colour Photography for the Undergraduate Laboratory', *American Journal of Physics* **38**, (11) (1970).

11. R. Child Bayley, *Photography in Colours*, Second Edition, Iliffe & Sons Limited, London,(1904).

12. N.J. Phillips, H. Heyworth, T. Hare, 'On Lippmann's Photography', *The Journal of Photographic Science* **32** (1984).

13. H.I. Bjelkhagen, 'Lippmann photographs recorded in DuPont color photopolymer material', *Proc. SPIE* **3011**, 358-366 (1997).

14. Jean-Marc Fournier, 'An investigation on Lippmann Photographs: Materials, Processes and Color Rendition', *Proc. SPIE* **2176**, (1994).

15. Helge Nareid, 'A Review of the Lippmann Colour Process', *The Journal of Photographic Science,* **36** (1988).

Appendix D

Computer simulation of the Lippmann photographic process and recording experiments using holographic materials (Paper)

Computer simulation of the Lippmann photographic process and recording experiments using holographic materials

Hans I. Bjelkhagen* and Savita De Souza

De Montfort University, The Centre for Modern Optics
Hawthorn Building, The Gateway, Leicester LE1 9BH
United Kingdom

ABSTRACT

The old Lippmann color photography technique (invented 1891) has been investigated. Today, high-resolution panchromatic recording materials are on the market suitable for Lippmann photography. The holographic panchromatic silver-halide materials from Slavich in Russia, as well as the panchromatic photopolymers from Du Pont in the USA can be used for recording Lippmann photographs. In particular, the Slavich emulsion has been investigated for recording Lippmann photographs. In order to understand the Lippmann technique better, computer simulations of the recording and reconstruction process have been undertaken. A Lippmann color photograph is unique and almost impossible to copy, which makes it suitable for the document security application. A new Optical Variable Device (OVD) technique is under development. This type of photograph can be applied to personal documents as a new optical security device, e.g., passports, ID cards, etc., can carry a Lippmann OVD. The recording of a Lippmann photograph requires a special panchromatic recording material, e.g., silver halide materials or photopolymer materials.

Keywords: Color photography, Lippmann photography, interferential photography, optical security, optical variable devices, photopolymer materials, silver halide materials

1. INTRODUCTION

After the invention of black-and-white photography in the 19th century a lot of research was devoted to the possibility of recording natural color images. In 1886 Lippmann developed a general theory of recording colors as standing waves in a light-sensitive emulsion. However, most of his time was devoted to perfecting a suitable recording emulsion for his experiments. In 1891 he succeeded in recording first a spectrum and later color photographs. Lippmann was a awarded the Physics Nobel Prize in 1908 for his photographic technique known as *interferential photography* or *interference color photography.* His early work was published in several papers[1–4]. In this type of photography, color is recorded in a photosensitive film as a black-and-white interference structure. The photographic image in a Lippmann photograph is Bragg sensitive, which means it is suitable for the optical document security application. The preserved 100-year-old Lippmann photographs are very beautiful, having extremely high resolution and good color contrast. There are several recent publications on this interesting imaging technique[5–16].

Optical Variable Devices (OVDs), such as holograms, are now common in the field of document and product security[17]. The Lippmann photograph represents a new type of OVD, which belongs to the interference security image structures (ISISs). It offers additional advantages over embossed hologram for unique document security applications. The main advantage of this new device is that it can be individually made, which means, a unique label for each security document such as passports, identification cards, etc. This particular application has been introduced by Bjelkhagen[18–20].

*Correspondence: Email: hansholo@aol.com; Telephone: +44-116 250 6374; Fax: +44-116 250 6144

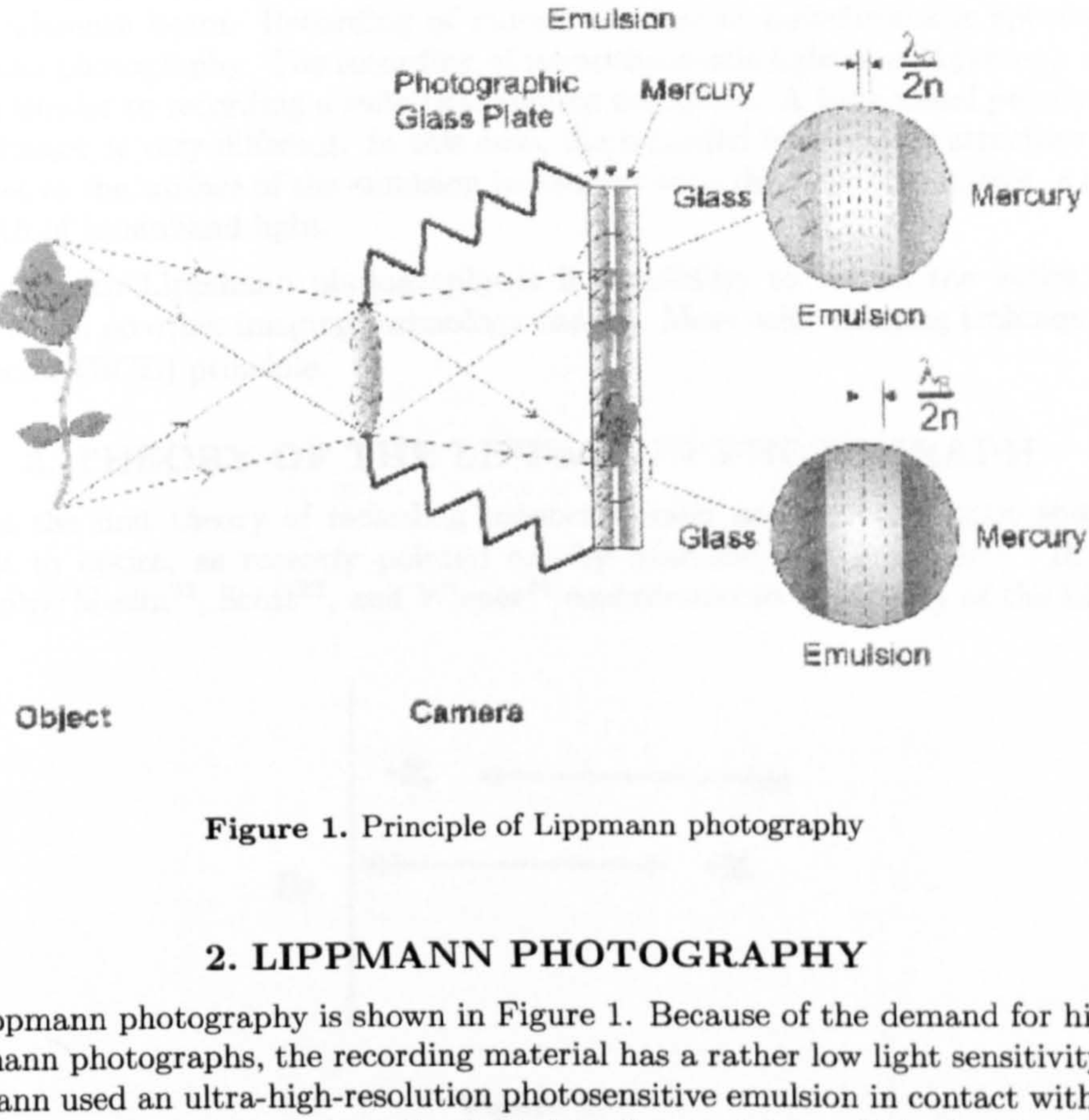


Figure 1. Principle of Lippmann photography

2. LIPPMANN PHOTOGRAPHY

The principle of Lippmann photography is shown in Figure 1. Because of the demand for high resolving power for recording Lippmann photographs, the recording material has a rather low light sensitivity. To record such a photograph, Lippmann used an ultra-high-resolution photosensitive emulsion in contact with a highly reflecting surface such as mercury. This mirror reflects the light into the emulsion, which then interferes with the light coming from the other side of the emulsion. Stationary standing waves of the interfering light produce a very fine fringe pattern throughout the emulsion with a periodic spacing of $\lambda/(2n)$ that is recorded (λ is the wavelength of light in air and n is the refractive index of the emulsion). The color information concerning the object is recorded in this way. For example, a large separation between the fringes in the emulsion indicates that the recorded wavelength is located at the red end of the spectrum. More closely spaced fringes indicate a shorter wavelength, such as green or blue. This description is correct only when rather monochromatic colors are recorded.

When the developed photograph is viewed in white light, different parts of the recorded image produce different colors due to the separation of the recorded fringes in the emulsion. The light is reflected from the fringes, creating different colors corresponding to the original ones that produced them during the recording. It is obvious that there is a severe requirement on the resolving power to record the fringes separated in the order of half the wavelength of the light. It is also clear that the processing of these plates is critically important, as one can not change the separation between the fringes since that would create wrong colors. To observe the correct colors, the illumination and observation must be at normal incidence. If the angle changes, the color of the image will change. This change of color, known as iridescence, is a very important feature of the Lippmann photograph as a document security device. The Lippmann image is recorded as a Bragg structure in the emulsion.

Modern holography shows similarities to Lippmann photography. In both cases an interference pattern is recorded in a high-resolution emulsion. The Bragg diffraction regime applies to both categories. The fundamental difference is that, in the Lippmann case, there is no phase recording involved; the recorded interference structure is a result of phase-locking the light by the reflecting mirror. In holography, the phase information

is actually recorded, being encoded as an interference pattern created between the light reflected from the object and a coherent reference beam. Recording of monochromatic or polychromatic spectra must be treated differently in Lippmann photography. The recording of monochromatic light in a Lippmann emulsion is easy to understand, and it is similar to recording a reflection volume hologram. A broadband polychromatic spectrum, such as a landscape image, is very different. In this case, the recorded interference structure in the emulsion is located only very close to the surface of the emulsion in contact with the reflecting mirror, a consequence of the short coherence length of broadband light.

What is so interesting with Lippmann photography is it possibility to record the entire spectrum of light reflected off objects, which no other imaging technology can do. Most color imaging techniques today are based on Maxwell's three color (RGB) principle.

3. THEORY OF THE LIPPMANN PHOTOGRAPH

Lippmann developed the first theory of recording monochromatic and polychromatic spectra[4]. He applied Fourier mathematics to optics, as recently pointed out by Marraud and Fournier[21]. In the early days of Lippmann photography, Meslin[22], Schtt[23], and Wiener[24] contributed to the theory of the Lippmann process.

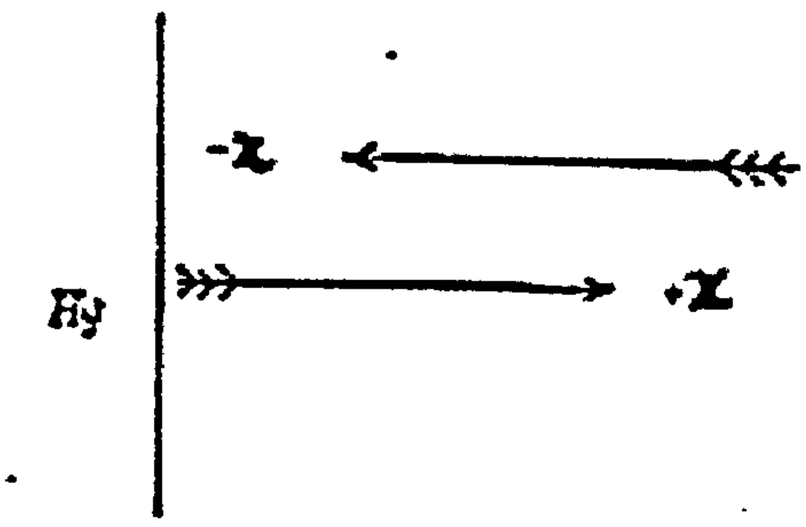


Figure 2.

In Figure 2, if -x is the direction of light coming from the object, and x is the direction of the reflected light, we have the intensity of the interference pattern as a function of x, i.e. the emulsion is exposed to a standing wave. This is shown in Equ.1 below:

$$i = c \bullet \sin^2 \frac{2\pi x}{\lambda} \tag{1}$$

If we assume that we have no absorption or scattering taking place in the emulsion, the density variation in the emulsion will be proportional to the intensity distribution as illustrated in Figure 3.

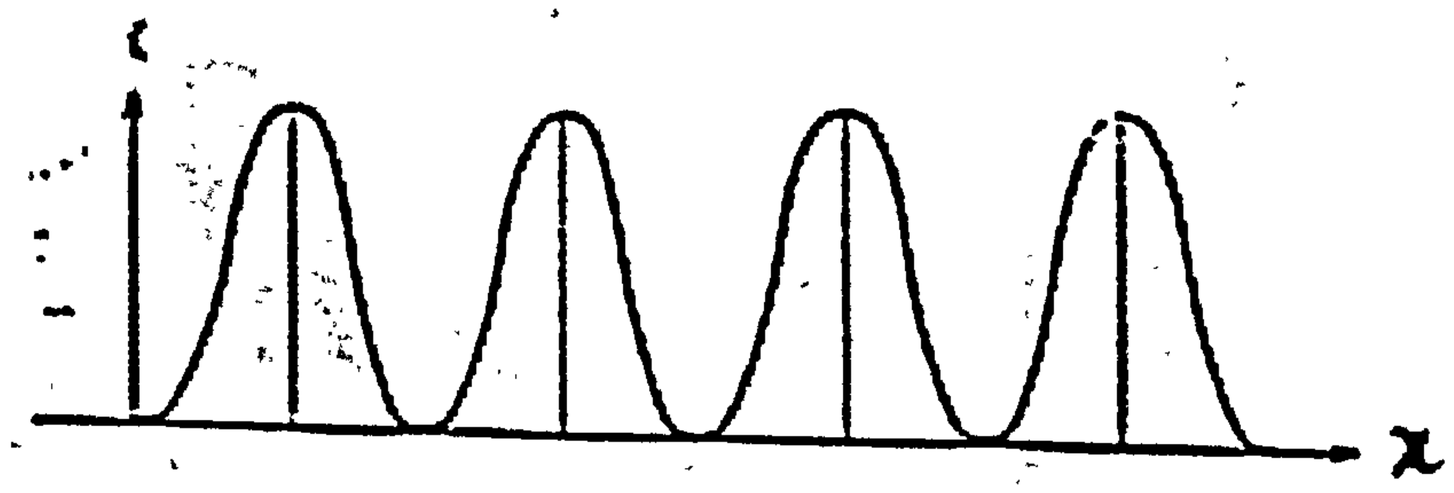


Figure 3.

If we consider the case when we have the simultaneous recording of two different wavelengths in an emulsion, the resulting intensity distribution is given by:

$$i' = i_1 + i_2 = c \bullet \left(\sin^2 \frac{2\pi x}{\lambda_1} + \sin^2 \frac{2\pi x}{\lambda_2} \right) \tag{2}$$

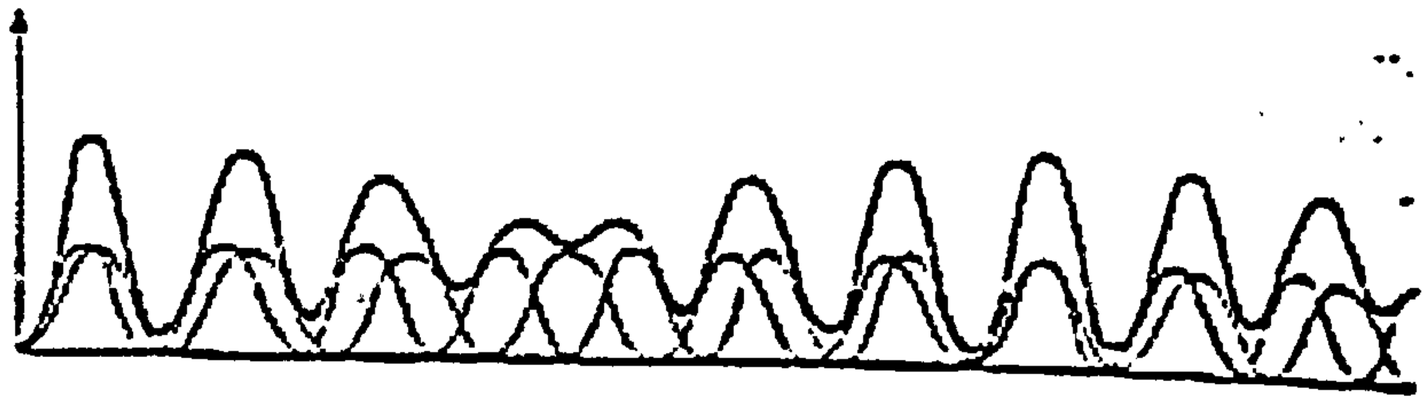


Figure 4.

In Figure 4, $\lambda_1 = 563$nm and $\lambda_2 = 482$nm.

In the case of recording white light in a Lippmann emulsion, we need to integrate the intensity variation over the whole visible spectrum. Hence Equ. 1 has to be integrated between λ_{red} to λ_{violet}.

$$\mathcal{F} = \int S(\lambda) \bullet F(\lambda) \bullet E(\lambda) \bullet \sin^2 \frac{2\pi x}{\lambda} d\lambda = c \bullet \int_{\lambda_{viol}}^{\lambda^{rot}} \sin^2 \frac{2\pi x}{\lambda} d\lambda \tag{3}$$

where $c = S(\lambda) \bullet F(\lambda) \bullet E(\lambda)$;

$S(\lambda)$ is the intensity of the light source as a function of the wavelength

$F(\lambda)$ is the portion of the reflected light from the object as a function of the wavelength

$E(\lambda)$ is the sensitivity of the photographic emulsion as a function of the wavelength

For a Lippmann emulsion, we assume an isochromatic plate and hence $E(\lambda)$ is constant.

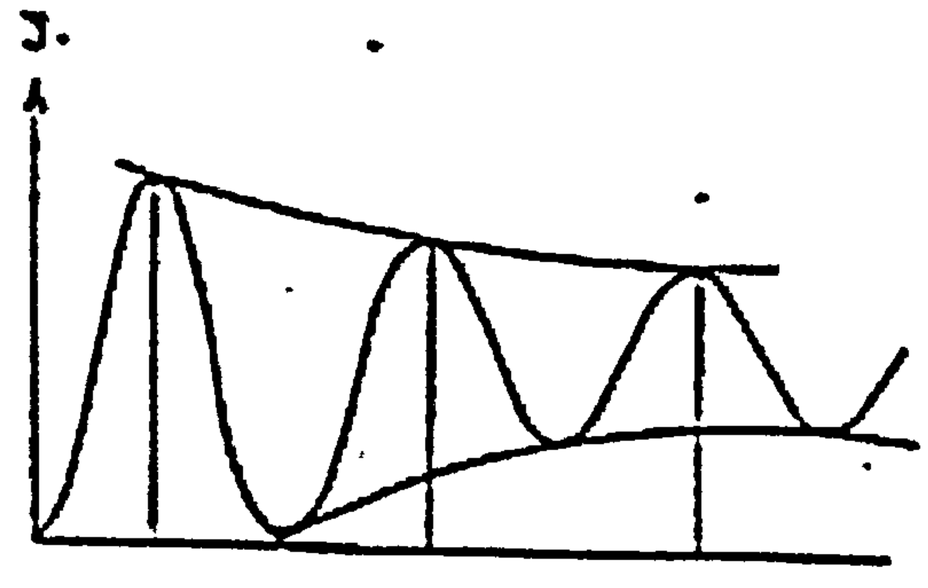


Figure 5.

If $S(\lambda)$, $F(\lambda)$, $E(\lambda)$ are constant, then c is a constant and hence can be taken out of the integral sign. The graphical representation is shown in Figure 5.

The above theory is the most simplistic approach but it gives a feeling of how the Lippmann process works. In reality, the recording material is not free of absorption and scattering and these properties have been taken into consideration by Phillips et al[27]. These factors will be considered in future work using the Light Tools simulation software.

Lehmann[25], a German photo scientist and early practitioner of Lippmann photography, performed recording experiments to verify that the theoretically predicted interference patterns actually were recorded in the emulsion. He used optical microscopy to record microphotographs of emulsion samples in which three different recordings had been performed. The predicted interference patterns described above were recorded in the emulsion as shown in Figures 6 - 8. The magnification of the emulsion is between 7,000 and 10,000 times in these photographs recorded by Lehmann[25].

Figure 6. Monochrome red light ($\lambda = 625$ nm)

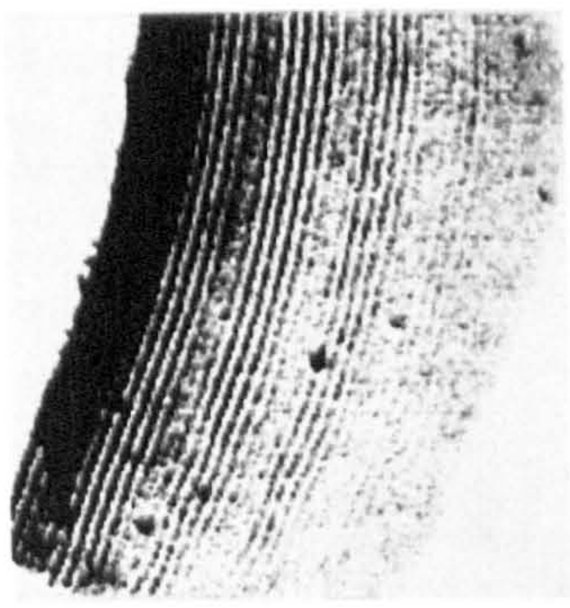


Figure 7. Two wavelengths recording ($\lambda = 482$ nm and 563 nm)

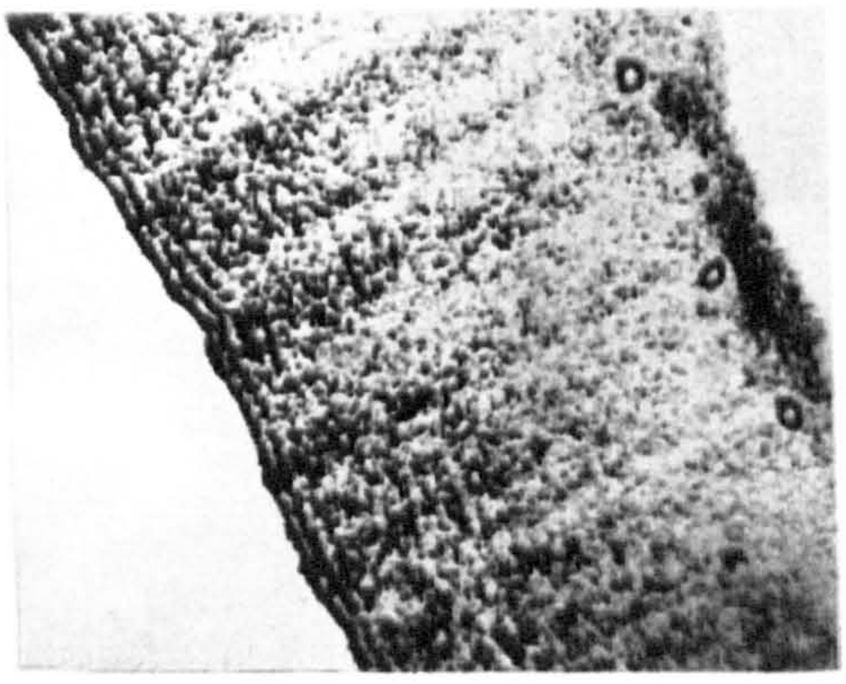


Figure 8. Broadband recording (Peaked at about $\lambda = 575$ nm)

Fournier and Burnett[26] investigated old Lippmann photographs using modern electron microscopy, where similar interference structures recorded in the emulsion were revealed, although with a much better resolution. In addition, they described similarities between Lippmann photography and holography. Another important contribution was made by Phillips et al.[26] comparing the theory of Lippmann photography and holography.

4. SIMULATION OF THE LIPPMANN PROCESS

The first researchers to perform computer simulation of the Lippmann process were Nareid and Pedersen.[28,29] For monochromatic recording, Kogelnik's coupled-wave theory[30] can be applied to Lippmann photography. For polychromatic recording, superposition of several frequencies in the recording light gives rise to aperiodic space gratings for which the Kogelnik theory cannot be applied. Instead, Nareid and Pedersen based their modeling on the theory of wave propagation in a stratified medium combined with first Born approximation.

4.1. Simulation of the Lippmann process

It has been determined that the calculations for the simulation process are one-dimensional calculations. The transmission line model is a very powerful method for analysing 1D wave propagation in stratified media. Mathematically it is equivalent to the transfer matrix method but instead of 2x2 matrices one uses recursion relations, which yield simpler algorithms, especially for reflection coefficient calculations.

4.2. Reflection coefficient by the transmission line model

We first show how the plane wave response of a vertically stratified medium can be computed by use of the transmission line model. We only consider plane waves in the $\pm z$ direction but the method is readily generalized to arbitrary incident fields by means of a Fourier expansion of the x,y dependence (the angular spectrum expansion).

We assume the stratified medium to consist of N layers of finite thickness on top of an infinitely thick bottom layer. The m'th layer ($m = 0, 1, 2...N$) has thickness d_m ($d_N \to \infty$), a refractive index $n_m = \sqrt{\epsilon_m/\epsilon_0}$, and a wave impedance $\eta_m = \eta_0/\eta_m$, where $\eta_0 = \sqrt{\mu_0/\epsilon_0} \simeq 120\pi\Omega$ is the free space wave impedance. At any point

4.3. The first Born approximation

The Equ. 6 shows that ρ_0 is generally a non-linear function of the parameters of the medium. However, a linear model is obtained in the weak scattering or first Born approximation. It applies for $|\rho_m^0|$, $|\rho_m| << 1$ in which case we can use the approximation

$$\frac{\rho_m^0 + \rho_{m+1}}{1 + \rho_m^0 \rho_{m+1}} \cong \rho_m^0 + \rho_{m+1}, \tag{9}$$

in Equ. 6. We then obtain

$$\rho_m \cong (\rho_m^0 + \rho_{m+1}) \exp(2ikn_m d_m), \tag{10}$$

and the iteration becomes trivial, yielding:

$$\rho_0^B = \sum_{m=0}^{N-1} \rho_m^0 \exp(2ik \sum_{s=0}^{m} n_s d_s), \tag{11}$$

where the superscript "B" is used to indicate the first Born approximation.

In the low frequency limit $k \to 0$, we now obtain

$$\rho_0^B \to \sum_{m=0}^{N-1} \rho_m^0 \neq \frac{\eta_N - \eta_0}{\eta_N + \eta_0} \tag{12}$$

which, however, is not consistent with Equ. 8. This inconsistency is removed when the Fresnel coefficients in Equ. 7 are approximated by:

$$\rho_m^0 = \frac{\eta_{m+1} - \eta_m}{\eta_{m+1} + \eta_m} \cong \frac{1}{2} \ln(\eta_{m+1}/\eta_m) \tag{13}$$

Then Equ. 12 yields

$$\rho_0^B \to \sum_{m=0}^{N-1} \rho_m^0 = \frac{1}{2} \ln(\eta_N/\eta_0) \cong \frac{\eta_N - \eta_0}{\eta_N + \eta_0} \tag{14}$$

The consistent first Born approximation is therefore given by:

$$\begin{aligned} \rho_0^B &= \sum_{m=0}^{N-1} \frac{1}{2} \ln(\eta_{m+1}/\eta_m) \exp(2ik \sum_{s=0}^{m} n_s d_s) \\ &= -\sum_{m=0}^{N-1} \frac{1}{2} \ln(\eta_{m+1}/\eta_m) \exp(2ik \sum_{s=0}^{m} n_s d_s) \end{aligned} \tag{15}$$

In the continuous limit ($d_m \to d_x$, $N \to \infty$) we now obtain the usual formula:

$$\rho_0^B(k) = -\frac{1}{2} \int_0^\infty dz \frac{d \log n(z)}{dz} \exp(2ik \int_0^z n(z')dz') \tag{16}$$

At this point we introduce the two-way optical path-length:

$$u = u(z) = 2\int_0^z n(z')dz' \tag{17}$$

as a new integration variable. We then obtain:

$$\begin{aligned} \rho_0^B(k) &= -\frac{1}{2}\int_0^\infty du \frac{d\ln n(u)}{du}\exp(iku) \\ &= \frac{1}{2}ik\int_0^\infty du \ln[n(u)]\exp(iku) \end{aligned} \tag{18}$$

where the last expression follows from an integration by parts and the fact that we have assumed $n(0) = 1$.

5. RECORDING MATERIALS FOR LIPPMANN PHOTOGRAPHY

New and improved recording materials combined with special recording and processing techniques have made it possible to develop the new OVD method. In particular, the ultra-high image resolution, the Bragg sensitivity of the image, the archival image color stability, the difficulty in copying such photographs, and the possibility to individually record them, make them suitable for the document security application. The optimal recording material for Lippmann photography has to be isochromatic to give a correct color recording. Most holographic materials are not isochromatic and therefore color correction filters may be needed for the recording of Lippmann photographs.

There are two holographic recording materials suitable for Lippmann photography: panchromatic ultra-high-resolution silver-halide materials and panchromatic photopolymer materials. Silver-halide materials requires wet processing. As regards such materials, only the new ultra-fine-grain panchromatic emulsions from Slavich can be considered[31,32]. In our case, the PFG-03C emulsion coated on film has been used. The grain size is between 10 nm and 20 nm for this emulsion. Some characteristics of the Slavich PFG-03C material are presented in Table 1.

Silver halide material	PFG-03C
Emulsion thickness	7 nm
Grain size	10 - 20 nm
Resolution	$\sim 10000 lp/mm$
Blue sensitivity	$\sim 1.0 - 1.5 10^{-3} J/cm^2$
Green sensitivity	$\sim 1.2 - 1.6 10^{-3} J/cm^2$
Red sensitivity	$\sim 0.8 - 1.2 10^{-3} J/cm^2$
Color sensitivity peaked at:	633 nm, 530 nm, 450 nm

Table 1. Characteristics of the Slavich PFG-03C emulsion

The panchromatic holographic photopolymer materials from E. I. du Pont de Nemours & Co.[33] are also suitable for Lippmann photography. They requires only dry processing (UV-curing and baking). This fact makes them particularly suitable for Lippmann OVD security products. Although, less sensitive than the Slavich silver-halide emulsion (which is also slow, according to modern photographic standards), it has its special advantages of easy handling and dry processing. However, a special type of photopolymer material is required for Lippmann photography. Successful recordings of Lippmann photographs in thin experimental photopolymer materials were reported by Bjelkhagen.[9]

A Lippmann photograph can be recorded in photopolymer materials in the following way. The photosensitive polymer layer has to be rather thin, in the order of a few micrometers only. The photopolymer layer must be

coated on a flexible transparent base and a special type of reflecting foil has to be laminated on top of to the photosensitive polymer layer in absolute perfect contact with it. The panchromatic photopolymer material having an emulsion thickness of about 2 to 4 μm is preferred. Such experimental materials have been manufactured by DuPont, e.g., the HRF-700X071-3 film. This particular polymer batch number was used for the first successful Lippmann photographs.[9] Before the recording takes place, a reflecting mirror foil has to be laminated to the polymer film. As a reflecting surface, silver sputtered (800Å) polyester film without the standard anti-dust oxide (InO) top layer was used. The mirror foil is laminated to the photopolymer material under safelight. A color correction filter may be needed in front of the camera lens to obtain correct color balance, since the photopolymer materials are not very isochromatic.

6. SECURITY APPLICATIONS OF LIPPMANN PHOTOGRAPHY

The possibility to use Lippmann photography for document security and counterfeit-resistant purposes was described in earlier publications.[18–20] For example, the application of a Lippmann photograph on security documents can be performed in the following way. The first step is the actual exposure of the document information (e.g., printed text, codes, signature, color photograph, etc.) to the recording film, similar to conventional photography, using a special macro lens. A special recording-processing device must be designed and manufactured in which a roll of the photosensitive film, prelaminated with a reflecting foil, is exposed. Special illuminating lamps, such as strong halogen spotlights are required to record the security documents with a reasonably short exposure time. After being exposed, the film must be processed. An automatic processing machine is required for this purpose. Actually, both the recording and the processing can be performed in the same piece of equipment. Because of the dry processing technique, the DuPont photopolymer materials would be the most suitable for an automatic Lippmann OVD system.

After being exposed to the image-forming information in the machine, the reflecting foil is detached from the photopolymer film and the photopolymer layer is exposed to strong white light or UV light. After that, the film has to pass through a hot section of the machine to increase the brightness of the image. The principle of such a system is shown in Figure 9.

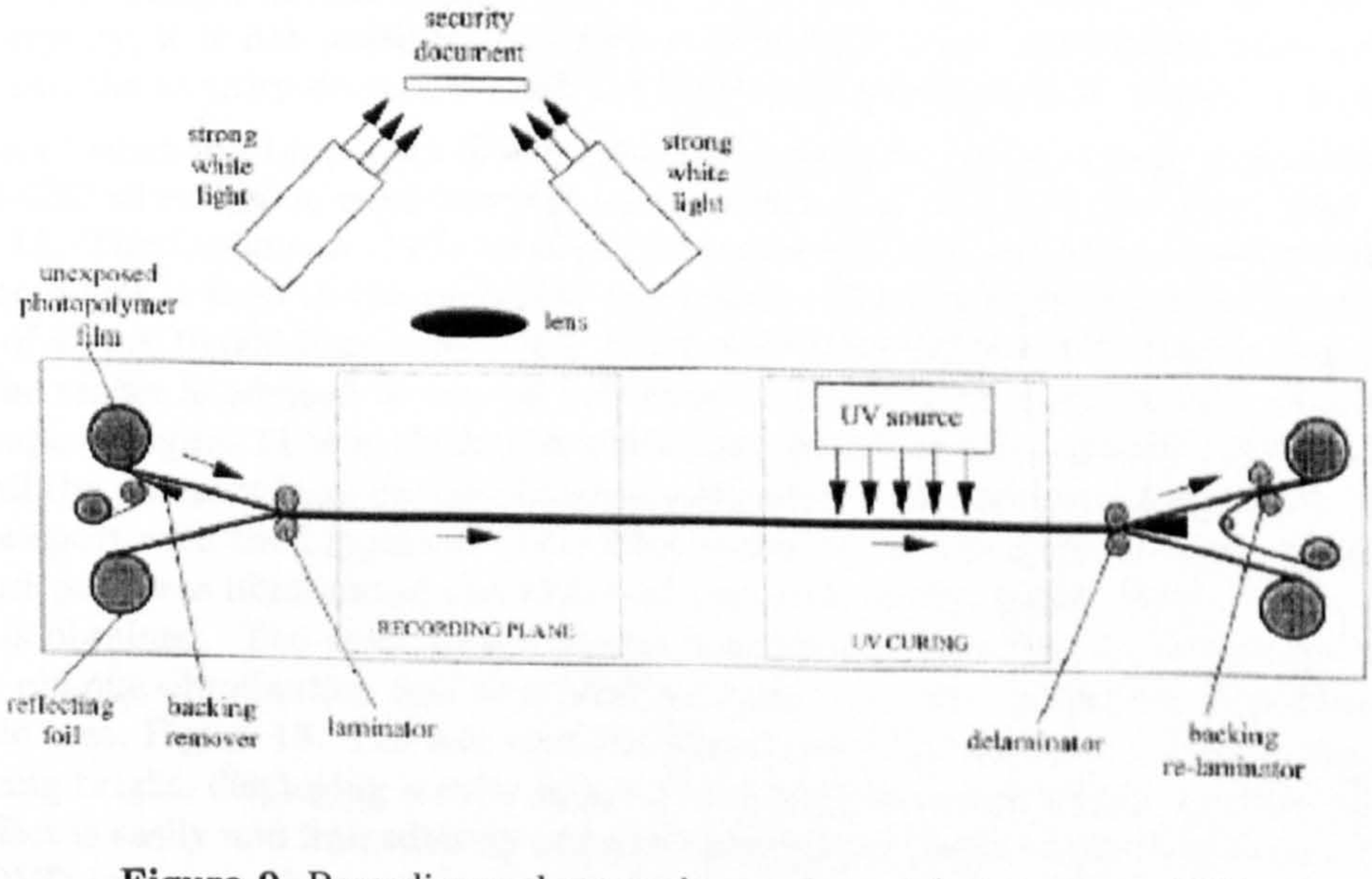


Figure 9. Recording and processing equipment for Lippmann OVDs.

After being processed the recorded images can be laminated to a black backing plastic foil. Then the photopolymer OVD is laminated to its corresponding security document. In the following a sample of a passport with a Lippmann OVD is presented.

Figure 10. Lippmann OVD attached to a security document.

Referring to Figure 10, in which a Lippmann OVD (1) is attached to a security document (2), the color of the image in the Lippmann OVD varies depending on the angle of observation. Perpendicular observation (3) gives the correct color image, oblique observation (4) shifts the colors toward shorter wavelengths. Because of the Bragg sensitivity, it is not possible to replace a Lippmann color photograph with a conventional color photograph, nor can the security document with the Lippmann photograph be copied in a color copier.

As an illustration of what the Lippmann OVD looks like, a sample passport page was recorded. In this case, the Slavich PFG-03C silver-halide emulsion was used instead of a photopolymer film. The results are shown in Figures 11 to 13. The Lippmann OVD itself has an extremely high resolution provided that a high-quality photographic macro-lens is used in the recording equipment. The recording silver halide emulsion itself has a resolving power of about 10,000 lines/mm. Only black-and-white digital photographic reproductions are used in this paper. The reader is advised to consult reference 18 to see color reproductions of Lippmann document security photographs. Figure 11 is a black-and-white photograph of the recorded sample passport Lippmann image in which all the passport information details are recorded with high resolution in the Lippmann OVD. In Figure 12, the passport with the Lippmann OVD label laminated to the passport page in the upper left corner. Then the passport page was illuminated and observed perpendicularly, which means that a correct color image of the passport is obtained. The color image change towards the blue part of the spectrum when the image is studied under oblique illumination and observation angles. At other angles the Lippmann OVD only looks like a dark plastic area, Figure 13. The fact that the Lippmann OVD switches between being completely black and suddenly being bright, displaying a color image of the passport page is very striking. Upon inspecting the passport, this effect is easily and immediately observed and very difficult to simulate in any other way. Further, the Lippmann OVD image can be compared with the passport page itself. The only tool needed for this may be a magnifying glass in order to be able to see and read the text in the high-resolution image.

Figure 11. B/W reproduction of a Lippmann photograph (Lippmann OVD) of the passport page.

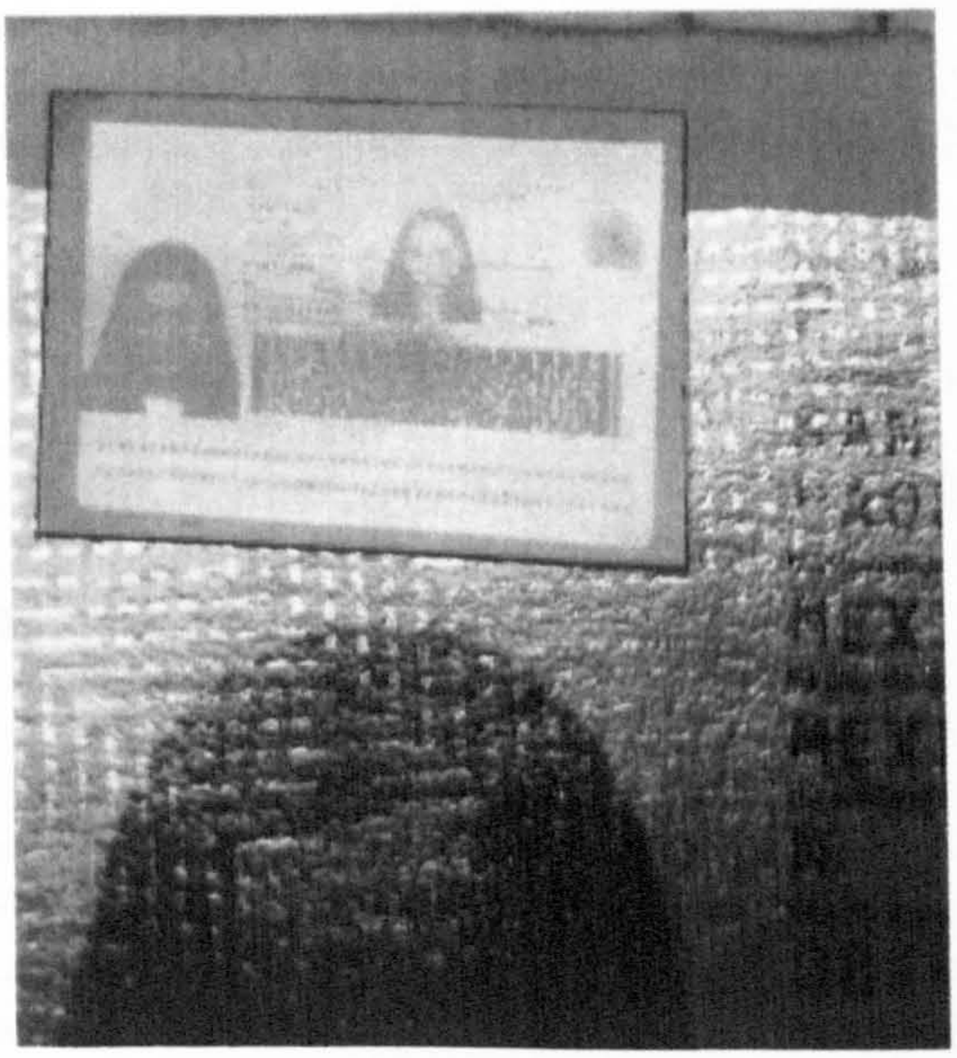


Figure 12. Passport page with a Lippmann OVD when the OVD is correctly illuminated and observed.

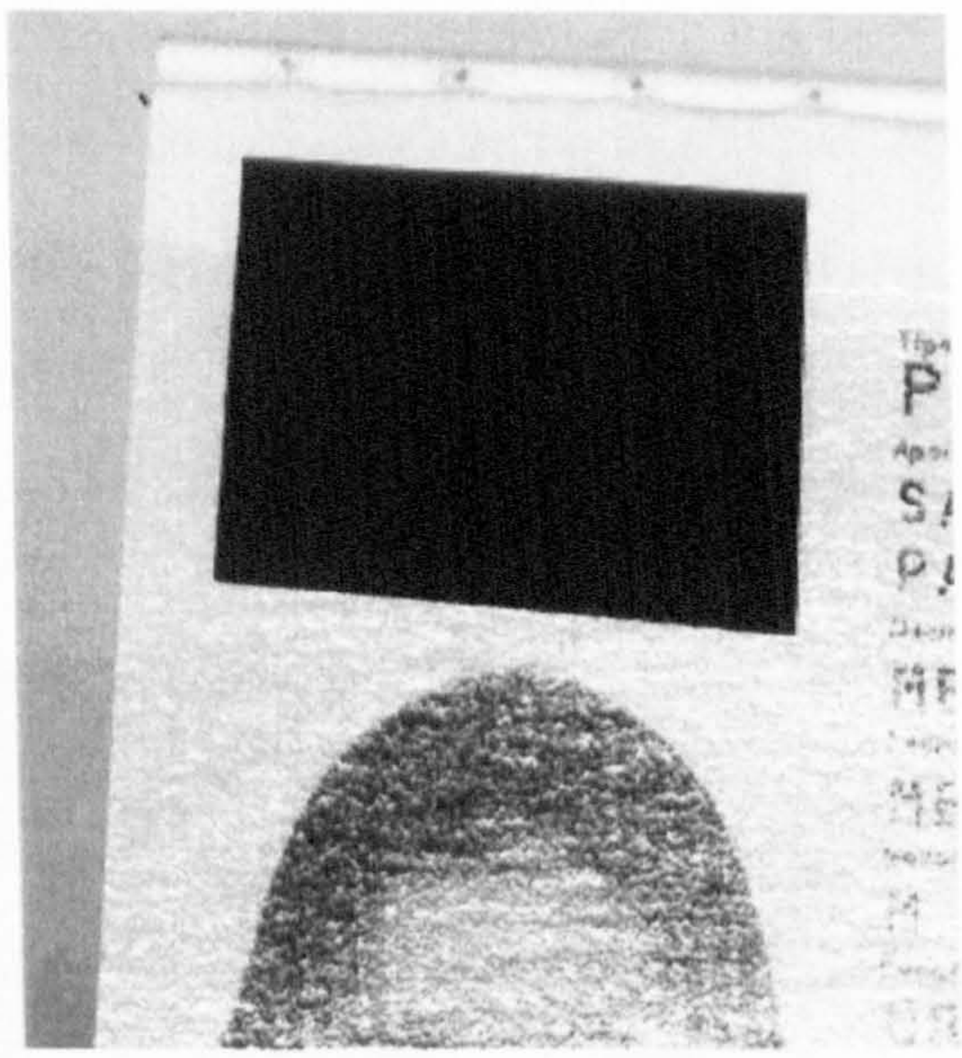


Figure 13. The passport page with the Lippmann OVD not visible.

REFERENCES

1. G. Lippmann, "La photographie des couleurs," *C. R. Hebd. Séances Acad. Sci.* **112**, pp. 274-275, 1891.
2. G. Lippmann, "La photographie des couleurs [deuxième note]," *C. R. Hebd. Séances Acad. Sci.* **114**, pp. 961-962, 1892.
3. G. Lippmann, "Photographies colorées du spectre, sur albumine et sur gélatine bichromatées," *C. R. Hebd. Séances Acad. Sci.* **115**, p. 575, 1892.
4. G. Lippmann, "Sur la théorie de la photographie des couleurs simples et composées par la méthode interférentielle," *J. Physique* **3** (3), pp. 97-107, 1894.

5. P. Connes, "Silver salts and standing waves: the history of interference color photography," *J. Optics (Paris)* **18**, pp. 147-166, 1987.
6. H. Nareid, "A review of the Lippmann color process," *J. Photogr. Sci.* **36**, pp. 140-147, 1988.
7. J.-M. Fournier, "Le photographie en couleur de type Lippmann: cent ans de physique et de technologie," *J. Optics (Paris)* **22**, pp. 259-266, 1991.
8. Yu.N Denisyuk, "Imaging properties of light intensity waves: the development of the initial Lippmann ideas," *J. Optics (Paris)* **22**, pp. 275-280, 1991.
9. C.C. Rich, L. Dickerson, "Lippmann photographic process put to practice with available materials," in *Holographic Materials II*, ed. by T.J. Trout. Proc. SPIE **2688**, pp. 88-95, 1996.
10. H.I. Bjelkhagen, "Lippmann photographs recorded in DuPont color photopolymer material," in *Practical Holography XI and Holographic Materials III*, ed. by S.A. Benton, T.J. Trout. Proc. SPIE **3011**, pp. 358-366, 1997.
11. Yu.N Denisyuk, "From Lippmann photography to selectograms via white light holography," *J. Imaging Sci. Technol.* **41**, pp. 205-210, 1997.
12. H.I. Bjelkhagen, T.H.Jeong, and R.J. Ro, "Old and modern Lippmann photography," in *Sixth Int'l Symposium on Display Holography*, ed. by T.H. Jeong. Proc. SPIE **3358**, pp. 72-83, 1998.
13. W.R. Alschuler, "On the physical and visual state of 100 year old Lippmann color photographs," in *Sixth Int'l Symposium on Display Holography*, ed. by T.H. Jeong. Proc. SPIE **3358**, pp. 84-94, 1998.
14. J.-M. Fournier, B.J. Alexander, and P.L. Burnett, S.E. Stamper, "Recent developments in Lippmann photography," in *Sixth Int'l Symposium on Display Holography*, ed. by T.H. Jeong. Proc. SPIE **3358**, pp. 95-102, 1998.
15. P. Hariharan, "Lippmann photography or Lippmann holography?" *J. Mod. Optics* **45**, pp. 1759-1762, 1998.
16. H.I. Bjelkhagen, "Lippmann photography: reviving an early colour process," *History of Photography* **23**, pp. 274-280, 1999.
17. R. L. van Renesse, Ed., *Optical Document Security*, 2nd ed. Artech House, Boston, London, 1998.
18. H.I. Bjelkhagen, "New optical security device based on one-hundred-year-old photographic technique," *Opt. Eng.* **38**, pp. 55-61, 1999.
19. H.I. Bjelkhagen, "A new OVD based on interferential photography recorded in holographic materials," in *Holographic Materials V*, ed. by T.J. Trout. Proc. SPIE **3638**, pp. 87-95, 1999.
20. H.I. Bjelkhagen, "The Lippmann OVD for enhanced document security," in *Optical Security and Counterfeit Deterrence Techniques III*, ed. by R.L. van Renesse, W.A. Vliegenthart. Proc. SPIE **3973**, pp. 276-283, 2000.
21. A. Marraud, J.-M. Fournier: Formation d'image accompagneé d'une analyse spectrale en champ complet: de J. Fourier à G. Lippmann. *Microsc. Microanal. Microstruct.* **8**, pp. 37-39, 1997.
22. G. Meslin: Sur la photographie des couleurs. *Annales de Chemie et de Physique* **27** (Ser.6), pp. 369-391, 1892.
23. F. Schütt: Innerer Bau und optisches Verhalten der Lippmannschen Photographien in natürlichen Farben. *Annalen der Physik und Chemie* **57**, pp. 533-554, 1896.
24. O. Wiener: Ursache und Beseitigungen eines Fehlers bei der Lippmann'schen Farbenphotographie, zugleich ein Beitrag zu ihrer Theorie. *Annalen der Physik und Chemie* **69**, pp. 488-530, 1899.
25. H. Lehmann, "Über die direkten Verfahren der Farbenphotographie nach Lippmann und Lumière," *Physikalische Zeitschrift* **8**, pp. 842-849, 1907.
26. J.-M. Fournier, and P. L. Burnett, "Color rendition and archival properties of Lippmann photographs," *J. Imaging Sci. Technol.* **38**, pp. 507-512, 1994.
27. N. J. Phillips, H. Heyworth, and T. Hare, "On Lippmann's photography," *J. Photogr. Sci.* **32**, pp. 158-169, 1984.
28. H. Nareid, and H. M. Pedersen, "Modelling of spectral response and tone reproduction in Lippmann photography and reflection holography," in *Practical Holography IV*, ed. by S.A. Benton. Proc. SPIE **1212**, pp. 63-72, 1990.

29. H. Nareid, and H. M. Pedersen, "Modeling of the Lippmann color process," *J. Opt. Soc. Am. A* **8**, pp. 257-265, 1991.
30. H. Kogelnik, "Coupled wave theory for thick hologram gratings," *Bell Sys. Tech. J.* **48**, pp. 2909-2947, 1969.
31. Yu. A. Sazonov, and P. I. Kumonko, "Holographic materials produced by the Micron plant at Slavich," in *Sixth Int'l Symposium on Display Holography*, T. H. Jeong, ed. Proc. SPIE **3358**, pp.31-40, 1998.
32. S. J. Zacharovas, D. B. Ratcliffe, G. R. Skokov, S. P. Vorobiov, P. I. Kumonko, and Yu. A. Sazonov, "Recent advances in holographic materials from Slavich," in *HOLOGRAPHY 2000*, T. H. Jeong, and W. K. Sobotka, eds. Proc. SPIE **4149**, pp.73-80, 2000.
33. S. H. Stevenson, "DuPont multicolor holographic recording film," in *Practical Holography XI and Holographic Materials III*, ed. by S. A. Benton, T. J. Trout. Proc. SPIE **3011**, pp. 231-241, 1997.

Bibliography

[1] Rudolf L. van Renesse. *Security Design of Valuable Documents and Products.* Artech House Publishers, 1998, ISBN:0-89006-982-4.

[2] Camus, Dubois, et al. *Security Papers and Special Effects.* Artech House Publishers, 1998, ISBN:0-89006-982-4.

[3] H.W. Nusmeier, J. Wotte. *Optical Security in Laminates.* Artech House Publishers, 1998, ISBN:0-89006-982-4.

[4] Rudolf L. van Renesse. *Iridescent Optically Variable Devices: A Survey.* Artech House Publishers, 1998, ISBN:0-89006-982-4.

[5] G. Colgate, Jr. *Document Protection by Holograms.* Artech House Publishers, 1998, ISBN:0-89006-982-4.

[6] J. A. Dobrowolski. *Optical thin film Security Devices.* Artech House Publishers, 1998, ISBN:0-89006-982-4.

[7] Rudolf L. van Renesse. *Liquid Crystal Security Devices.* Artech House Publishers, 1998, ISBN:0-89006-982-4.

[8] Michael Hendry, Pascal van Gimst. *Identification Cards.* Artech House Publishers, 1998, ISBN:0-89006-982-4.

[9] J.M. Haslop. *Security Printing Techniques.* Artech House Publishers, 1998, ISBN:0-89006-982-4.

[10] J.F.Moser. *Document Protection by Optically Variable Graphics (Kinegram)*. Artech House Publishers, 1998, ISBN:0-89006-982-4.

[11] M.T.Gale. *Zero-Order Grating Microstructures*. Artech House Publishers, 1998, ISBN:0-89006-982-4.

[12] Jack E.Cook. *Retroreflective Security Devices*. Artech House Publishers, 1998, ISBN:0-89006-982-4.

[13] Patrick Loo, "Digital Watermarking using Complex Wavelets," University of Cambridge, 2002.

[14] I.Pitas and T. Kaskalis. Applying Signature on Digital Images. *Proc. of IEEE Workshop on Non-linear Signal and Image Processing*. I. Pitas(Ed.). pp. 460-463.

[15] F. Mintzer, et.al. Effective and Ineffective Digital Watermarks. *IEEE Intl. Conference on Image Processing, ICIP-97*. Vol.3. pp.9-12.

[16] Fernando P'erez-Gonz'alez, Juan R. Hern'andez. *A Tutorial on Digital Watermarking*.

[17] R.G.V. Schyndel, A.Z. Tirkel and C.F.Osborne. A Digital Watermark. *IEEE Proceedings of International Conference in Image Processing*, volume 2, pages 86-90, 1994.

[18] K. Tanaka, Y. Nakamura and K. Matsui. Embedding Secret Information into a dithered multi-level image. *In MILCOM'90: A New Era. 1990 IEEE Military Communications Conference*, volume 1, pages 216-220, 1990.

[19] W. Bender, D. Gruhl, N. Morimoto, A. Lu. Techniques for data hiding. *IBM Systems Journal*, 35 (3&4):313-336, 1996.

[20] D.Berman, J.Bartell and D. Salesin. Multiresolution painting and compositing. *Computer Graphics (Proceedings of SIGGRAPH 94)*, pages 85-90, July 1994.

[21] J.S. De Bonet. Multiresolution sampling procedure for analysis and synthesis of texture images. *Computer Graphics (Proceedings of SIGGRAPH 97)*, pages 361-368, August 1997.

[22] L. Boney, A.H. Tewfik and K.N. Hamdy. Digital watermarks for audio signals. *In Proceedings of the third IEEE International Conference on Multimedia Computing Systems*, 1996, pages 473-480, 1996.

[23] A. Bruderlin and L. Williams. Motion Signal Processing. *In Computer Graphics (Proceedings of SIGGRAPH 95)*, pages 97-104, August 1995.

[24] A. Certain, J. Popovic, T. DeRose, T. Duchamp, D. Salesin and W. Stuetzle. Interactive multiresolution surface viewing. *Computer Graphics (Proceedings of SIGGRAPH 96)*, pages 97-115, August 1996.

[25] I.J. Cox, J.Kilian, F.T. Leighton and T. Shamoon. Secure spread spectrum watermarking for images, audio and video. *In IEEE Proceedings of International Conference on Image Processing*, volume 3, pages 243-246, 1996.

[26] J. Cox, J.Kilian, F.T. Leighton and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, 6(12):1673-1687, December 1997.

[27] J.J.K O'Ruanaidh, F.M. Boland, and O. Sinnen. Watermarking Digital Images for Copyright Protection. *In Proceedings of the Electronic Imaging and Visual Arts Conference*, pages 243-246, February 1996.

[28] J.J.K O'Ruanaidh, W.J. Dowling, and F.M. Boland. Phase watermarking of digital Images. *In IEEE Proceedings of International Conference on Image Processing*, pages 239-242, September 1996.

[29] E. Praun, H. Hoppe and A. Finkelstein. Robust mesh watermarking. *In Computer Graphics (Proceedings of SIGGRAPH 99)*, pages 49-56, August 1999.

[30] C.I.PodilChuk and W.Zeng. Image Adaptive Watermarking using Visual Models. *IEEE Journal on Selected Areas in Communications*, Vol.16, No.4, May 1998, pp.525-539.

[31] Ross Anderson. *Security Engineering: A Guide to Building Dependable Distributed Systems.* Wiley Computer Publishing, ISBN: 0-471-38922-6.

[32] F.A.P.Petitcolas, et al. Information Hiding - A Survey. *Proceedings of the IEEE*, Vol.87, No.7, July 1999, pp.1062-1078.

[33] F.A.P.Petitcolas, et al. Attacks on Copyright Marking Systems. *Proceedings of the Second International Workshop on Information Hiding (1998)*, Springer LNCS, v1525, pp 219-239.

[34] I.J. Cox, M.L. Miller and J.A. Bloom. *Digital Watermarking.* Academic Press, 2002, ISBN:1-55860-714-5.

[35] Eduardo Fullea and Jose M. Martinez. *Robust Digital Image Watermarking using DWT, DFT and Quality Based Average.*

[36] Erard, Kutter, et al. *How to Bypass the Wassenaar Arrangement: A New Application for Watermarking.*

[37] W.H. Press. *Numerical Recipes in C: The Art of Scientific Computing*, chapter 20, pages 896-901. Cambridge University Press, 1992.

[38] D. Kundur and D. Hatzinakos. Digital Watermarking for Telltale Tamper Proofing and Authentication. *Proceedings of the IEEE*, Vol. 87, pages 1167-1180,1999.

[39] C.S. Lu, H.Y. Mark Liao and C.J. Sze. Combined Watermarking for Image Authentication and Protection. *Proceedings of the First IEEE International Conference on Multimedia and Expo.*, USA, 2000.

[40] C.S. Lu, H.Y. Mark Liao and L.H. Chen. Multipurpose Audio Watermarking. *Proceedings of the 15th International Conference on Pattern Recognition*, Spain, 2000.

[41] G.J. Yu, C.S. Lu, H.Y. Mark Liao and J.P. Sheu. Mean Quantization Blind Watermarking for Image Authentication. *Proceedings of the 7th IEEE International Conference on Image Processing*, Canada, 2000.

[42] S. Bhattacharjee and M. Kutter. Compression Tolerant Image Authentication. *IEEE International Conference on Image Processing*, USA, 1998.

[43] J. Dittmann, A. Steinmetz and R. Steinmetz. Content-based Digital Signature for Motion Pictures Authentication and Content-Fragile Watermarking. *Proceedings of the IEEE Conference on Multimedia Computing and Systems*, vol. 11, Italy, 1999.

[44] C.Y. Lin and S.F. Chang. A Robust Image Authentication Method Surviving JPEG Lossy Compression. *SPIE Storage and Retrieval of Image/Videp Database*, Vol. 3312, San Jose 1998.

[45] C.Y. Lin and S.F. Chang. Generating Robust Digital Signature for Image/Video Authentication. *Multimedia Security Workshop at ACM Multimedia*, UK, 1998.

[46] Mark Liao, Lu, et al. *Structural Digital Signature for Image Authentication: An Incidental Distortion Resistant Scheme*, 2000.

[47] C.S. Lu, H.Y. Mark Liao, S.K. Huang and C.J. Sze. Cocktail Watermarking on Images. *Proceedings of the 3rd International Workshop on Information Hiding*, LNCS 1768, pages 333-347, September 29th - October 1st, 1999.

[48] R.B. Wolfgang and E.J. Delp. A Watermark for Digital Images. *Proceedings of the IEEE International Conference on Image Processing*, Sept. 16-19, 1996, Lausanne, Switzerland, pages 219-222.

[49] R. B. Wolfgang and E. J. Delp. A Watermarking Technique for Digital Imagery: Further Studies. *Proc. Intl. Conf. on Imaging Sciences, Systems and Tech.*, Las Vegas, June 30-Jul 3, 1997, pages 279-287.

[50] A.Z.Tirkel, et al. A Two-Dimensional Digital Watermark. *Proc. of Digital Image Computing, Technology and Applications, Brisbane, Australia*, 1995, pp.378-383.

[51] A.Z. Tirkel, G.A. Rankin, R.M. van Schyndel, et. al. Electronic Watermark. *DICTA-93 Macquarie University, Sydney*, December 1993, pages 666-672.

[52] S. Walton. Image Authentication for a Slippery New Age. *Dr.Dobb's Journal, April 1995*, pages 18-26, 82-87.

[53] Igar Djurovic, Srdjan Stankovic and I.Pitas. Digital Watermarking in the fractional Fourier Transformation domain. *Journal of Network and Computer Applications (2001)*, vol. 24, pages 167-173.

[54] Su, Wang and Kuo. Digital Watermarking on EBCOT Compressed Images. *SPIE's 44th Annual Meeting, Applications of Digital Image Processing XX11 (SD41)*, Denver, Colorado, July 18th - 23rd, 1999.

[55] Su, Kuo. An Image Watermarking Scheme to Resist Generalized Geometrical Transform. *SPIE Photonics East, Multimedia Systems and Applications 111 (VV08)*, Boston, November 5th -8th, 2000.

[56] Su, Kuo. Synchronised Detection of the Block-based Watermark with Invisible Grid Embedding. *SPIE Photonics West, Security and Watermarking of Multimedia Contents 111 (EI27)*, San Jose, California, Jan. 20th -26th, 2001.

[57] N. Terzija, M. Repges, K. Luck and W. Geisselhardt. *Impact of different Reed-Solomon Codes on Digital Watermarks based on DWT*, 2000.

[58] G. Mandyam and N. Ahmed. The Discrete Laguerre Transform: Derivation and Applications. *IEEE Transactions on Signal Processing*, vol. 44, No. 12, pages 2925-2931, December 1996.

[59] S.A.M. Gilani and A.N. Skodras. *DLT-Based Digital Image Watermarking.*

[60] J. Tian. Wavelet-based Reversible watermarking for Authentication. *Digimarc Corporation*, 19801 SW 72nd Avenue, Tualatin, OR97062, USA.

[61] G.S. Gulstad and K. Bruvold. *An Adaptive Digital Image Watermarking Technique for Copyright Protection.* ECE178, University of California, Santa Barbara.

[62] C. Wu and R. Cathey. *Digital Watermarking: A Comparative Overview of Several Digital Watermarking Schemes.* December 2002.

[63] Martin J. Turner et al. *Fractal Geometry in Digital Imaging.* Academic Press, 1998. ISBN:0-12-703970-8.

[64] Joseph W. Goodman. *Introduction to Fourier Optics.* Second Edition, McGraw Hill Publisher ISBN:0-07-024254-2

[65] Michael Barnsley. *Fractals Everywhere.* Academic Press ISBN: 0-12-079062-9

[66] Ning Lu. *Fractal Imaging.* Academic Press ISBN: 0-12-458010-6

[67] D.E. Knuth. *The Art of Computer Programming Volume 2: Seminumerical Algorithms.* Addison-Wesley, Reading, Mass, London 1969. ISBN:0-201-03822-6.

[68] William H. Press, Brian P. Flannery, Saul A. Teukolsky, William T. Vetterling. *Numerical Recipes, The Art of Scientific Computing.* Cambridge University Press ISBN: 0-521-30811-9

[69] M.M.Yeung. Digital Watermarking. *Communications of the ACM, Jul. 1998*, Vol.41, No.7, pp.31-33.

[70] N.Memon and P.W.Wong. Protecting Digital Media Content. *Communications of the ACM, July 1998*, Vol.41, No.7, pp.35-43.

[71] M.M Yeung, et al. Digital Watermarking for High-Quality Imaging. *IEEE First Workshop on Multimedia Signal Processing*, June23-25 1997, Princeton, New Jersey, pp. 357-362.

[72] C.Busch, et al. Digital Watermarking: From Concepts to Real-Time Video Applications. *IEEE Computer Graphics & Applications*, Jan/Feb 1999, pp.25-35.

[73] A.K.Jain. *Fundamentals of Digital Image Processing.* Prentice-Hall of India Pvt. Ltd., 1995.

[74] Rajmohan. *Watermarking of Digital Images.* ME Thesis Report, Dept. Electrical Engineering, Indian Institute of Science, Bangalore, India, 1998.

[75] S.P.Mohanty. *Watermarking of Digital Images.* Masters Project Report, Dept. of Electrical Engineering, Indian Institute of Science, Bangalore - 560 012, India, Jan 1999.

[76] B.Pfitzmann. Information Hiding Terminology. *Proc. of First Int. Workshop on Information Hiding*, Cambridge, UK, May30-Junel, 1996, Lecture notes in Computer Science, Vol.1174, Ross Anderson(Ed.), pp.347-350.

[77] W. Bendor, et. al. Techniques for Data Hiding. *IBM Systems Journal*, Vol.35, No.3 and 4, pp. 313-336, 1996.

[78] M.D.Swanson, et al. Transparent Robust Image Watermarking. *Proc IEEE International Conf. on Image Procesing, ICIP-96*, Vol.3, pp 211-214.

[79] C.T.Hsu and J.L.Wu. Hidden Digital Watermarks in Images. *IEEE Trans. on Image Processing*, Vol.8, No.1, Jan.1999, pp.58-68.

[80] P.Bas, et al. Using the Fractal Code to Watermark Images. *Proc. IEEE International Conf. on Image Processing, ICIP-98*, Vol.1, pp.469-473.

[81] J.Puate and F.Jordan. Using Fractal Compression Scheme to Embed a Digital Signature into an Image. *Proc. of SPIE Photonics East Symposuim*, Boston, USA, Nov. 18-22 1996.

[82] J. M. Acken. How Watermarking Adds Value to Digital Content? *Communications of the ACM*, July 1998, Vol.41, No.7, pp.75-77.

[83] J. Zhao, et. al. In Business Today and Tommorrow. *Communications of the ACM*, July 1998, Vol.41, No.7, pp.67-72.

[84] R.Barnett. Digital Watermarking : Application, Techniques, and Challenges. *IEE Electronics and Communication Engineering Journal*, August 1999, pp.173-183.

[85] J.S.Craver, et al. Can Invisible Watermarks Resolve Rightful Ownership? *IBM Research Report, RC205209*, July25 1996.

[86] M.Ramkumar and A.N.Akansu. Image Watermarks and Counterfeit Attacks : Some Problems and Solutions. *Proc. of Content Security and Data Hiding in Digital Media*, Newark, NJ, USA, May 14 1999.

[87] H. I. Bjelkhagen. Lippmann photographs recorded in DuPont colour photopolymer material. *In Practical Holography XI and Holographic Materials III, ed. by S.A. Benton, T. J. Trout. Proc. SPIE 3011*, pp. 358-366, 1997.

[88] Yu. N Denisyuk. From Lippmann photography to selectograms via white light holography. *J. Imaging Sci. Technol. 41*, pp. 205-210, 1997.

[89] H. I. Bjelkhagen, T. H. Jeong, and R. J. Ro. Old and modern Lippmann photography. *In Sixth Int'l Symposium on Display Holography, ed. by T. H. Jeong. Proc. SPIE 3358*, pp. 72-83, 1998.

[90] W.R. Alschuler. On the physical and visual state of 100 year old Lippmann colour photographs. *In Sixth Int'l Symposium on Display Holography, ed. by T. H. Jeong. Proc. SPIE 3358*, pp. 84-94, 1998.

[91] J. M. Fournier, B.J. Alexander, P.L. Burnett, and S.E. Stamper. Recent developments in Lippmann Photography. *In Sixth Int'l Symposium on Display Holography, ed. by T. H. Jeong. Proc. SPIE 3358*, pp. 95-102, 1998.

[92] P. Hariharan. Lippmann photography or Lippmann holography? *J. Mod. Optics 45*, pp. 1759-1762, 1998.

[93] H. I. Bjelkhagen. Lippmann photography: reviving an early colour process. *History of Photography 23*, pp. 274-280, 1999.

[94] H. I. Bjelkhagen. New optical security device based on one-hundred-year-old photographic technique. *Opt. Eng. 38*, pp. 55-61, 1999.

[95] P. Connes. Silver salts and standing waves: the history of interference colour photography. *J. Optics (Paris) 18*, pp. 147-166, 1987.

[96] J. Belloni et al. Enhanced yield of photoinduced electrons in doped silver halide crystals. *Letters to Nature*, vol. 402, 23/30 December 1999.

[97] H. Nareid. A review of the Lippmann colour process. *J. Photogr. Sci. 36*, pp. 140-147, 1988.

[98] Yu. N Denisyuk. Imaging properties of light intensity waves: the development of the initial Lippmann ideas. *J. Optics (Paris) 22*, pp. 275-280, 1991.

[99] C.C. Rich, L. Dickerson. Lippmann photographic process put to practice with available materials. *In Holographic Materials II, ed. By T. J. Trout. Proc. SPIE 2688*, pages 88-95, 1996.

[100] Anni Berger-Schunn. *Practical Color Measurement* John Wiley & Sons Inc. ISBN: 0-471-00417-0.

[101] Dr. R.W.G.Hunt. *Measuring Colour.* Third Edition, Fountain Press, ISBN:0-86343-387-1.

[102] H. I. Bjelkhagen. A new OVD based on interferential photography recorded in holographic materials. *In Holographic Materials V, ed. by T.J. Trout. Proc. SPIE 3638*, pp. 87-95, 1999.

[103] H. I. Bjelkhagen. The Lippmann OVD for enhanced document security. *In Optical Security and Counterfeit Deterrence Techniques III, ed. by R. L. van Renesse, W.A. Vliegenthart, Proc. SPIE 3973*, pp. 276-283, 2000.

[104] J.M. Fournier, P.L. Burnett. Colour rendition and archival properties of Lippmann photographs. *J. Imaging Sci. Technol. 38*, pp. 507-512, 1994.

[105] N.J. Phillips, H. Heyworth and T. Hare. On Lippmann's photography. *J. Photogr. Sci. 32*, pp. 158-169, 1984.

[106] H. Nareid and H. M. Pedersen. Modelling of spectral response and tone reproduction in Lippmann photography and reflection holography. *In Practical Holography IV, ed. by S.A. Benton, Proc. SPIE 1212*, PP. 63-72, 1990.

[107] H. Nareid and H. M. Pedersen. Modelling of the Lippmann colour process. *J. Opt. Soc. Am. A 8*, pp. 257-265, 1991.

[108] H. Kogelnik. Coupled wave theory for thick hologram gratings. *Bell Sys. Tech. J. 48*, pp. 2909-2947, 1969.

[109] S.J. Zacharovas, D.B. Ratcliffe, G.R. Skokov, S.P. Vorobiov, P.I. Kumonko and Yu. A. Sazonov. Recent advances in holographic materials from Slavich. *In HOLOGRAPHY 2000, T.H. Jeong, and W.K. Sobotka, eds. Proc. SPIE 4149*, pp. 73-80, 2000.

[110] S.H. Stevenson. DuPont multi-colour holographic recording film. *In Practical Holography XI and Holographic Materials III, ed. by S.A.Benton, T.J. Trout. Proc. SPIE 3011*, pp. 231-241, 1997.

[111] J.M. Blackledge. *Quantitative Coherent Imaging: Theory, Methods and Some Applications.* Academic Press. 1989. ISBN:0-12-103300-7

[112] Marc Levoy. The Digital Michelangelo Project. *Eurographics 1999.* Department of Computer Science, Stanford University.